2021

Design, development and usability evaluation of social system interface and development of computational model

Khoshnevis, Pedram

Design, Development and Usability Evaluation of Social System Interface and Development of Computational Model

by

Pedram Khoshnevis

A thesis submitted in partial fulfillment of the

requirements for the degree of

Master of Science

in

Computer Science

in the

Faculty of Science and Environmental Studies

of

Lakehead University, Thunder Bay

Committee in charge:

Dr. Vijay Mago (Principal Supervisor)
Dr. Eric Latimer (Co-supervisor)
Dr. Piper Jackson (External Examiner)
Dr. Muhammad Asaduzzaman (Internal Examiner)

Winter 2021

The thesis of Pedram Khoshnevis, titled Design, Development and Usability Evaluation of Social System Interface and Development of Computational Model, is approved:

Chair ______________________________ Date ____________

______________________________ Date ____________

______________________________ Date ____________

Lakehead University, Thunder Bay

Design, Development and Usability Evaluation of Social System Interface and Development of Computational Model

Abstract

DESIGN, DEVELOPMENT AND USABILITY EVALUATION OF SOCIAL SYSTEM INTERFACE AND DEVELOPMENT OF COMPUTATIONAL MODEL

In recent times, methods of computational intelligence (CI) that aim to solve real-life problems are developed by computer science researchers in collaboration with domain experts. There has also been an increased emphasis on the usability aspect of these algorithms by developing easy-to-use web interfaces. The graphical user interfaces (GUIs) designed for these algorithms are often designed solely to connect the web interfaces to the algorithm's functionality. While this is effective from researchers' perspective, the needs of new users (such as policymakers) in relation to software use are often neglected. The lack of consideration of new users' experience when developing GUIs often establishes usability issues for the technology and as a result expands the gap between the advances made in the computer science field and other fields, most notably the social sciences. This thesis investigates the various design, development, and evaluation methods for social simulation software and provides valuable insights for researchers and user interface designers who seek to create an effective GUI. Additionally, this thesis provides a case study of how computational models can be effectively applied for approaching complex social problems such as homelessness. In chapter 3 the development and testing process of the Homelessness Visualization (HOMVIZ) platform is discussed. The HOMVIZ platform uses a deep learning algorithm in order to predict potential trends in homeless populations in a particular area of interest. Various aspects of the user interface (UI) design were analyzed and a 14 participant usability testing session was conducted in order to discern the perceived usability of the platform. The UI

evaluation session in this chapter involved software testing, focus groups, and questionnaires. These sessions provided our research with valuable qualitative and quantitative data. Chapter 4 explores moderated and unmoderated usability testing sessions and compares them in terms of efficiency, reliability, and flexibility. The research for this chapter was approved by the Lakehead University's Research Ethics Board. The usability testing was conducted with a sample size of 72 participants. The research presented in this chapter provides valuable insight regarding different usability testing session methods and the impact of a known phenomenon called careless responding (CR) on data quality. Chapter 5 provides an example of how computational models can help mitigate a more complex social problem such as homelessness. The research presented in this chapter focuses on the operation of homeless shelters within Canada and introduces eight computation models that have the potential to improve the quality of life of people experiencing homelessness.

Dedication

This thesis work is dedicated to my family, friends, and all those who have helped me along my journey. To my father, mother, and sister for their continued support. To my aunt for her continued effort in bringing me comfort while I reside in North America. And finally, to my supervisor for his mentorship.

Contents

List of Figures

List of Tables

Acknowledgments

Foremost, I would like to express my sincere gratitude to my advisor, Dr. Vijay Mago, for his patience, motivation, enthusiasm, immense knowledge, and continuous support of my MSc studies and research. His guidance has helped me throughout my research and my writing of this thesis. I could not have asked for a better advisor and mentor for my MSc study.

My sincere thanks also go to Dr. Eric Latimer, Dr. Tim Aubry, Dr. Bartosz Gajderowicz, Savanah Tillberg and Dr. Abhi Rao for helping me with the technical and grammatical aspects of writing my thesis and research articles.

I thank my fellow labmates, colleagues, and other members of the DaTALab for their feedback on my writing and their continued encouragement. This research is supported by Insight grant-21820 from the Social Sciences and Humanities Research Council (SSHRC).

Lastly, I would like to thank my family: Azar Amigh Momayez, Hossein Khoshnevis, Pantea Khoshnevis, and Helena Khoshnevis for their continued support throughout my life.

Chapter 1

Introduction

1.1 Overview and Problem Statement

Complex social problems such as homelessness and poverty can greatly benefit from the latest developments in computational intelligence. The social sector provides life-sustaining aids to the most distressed and vulnerable population through the support from government and private organizations [1]. With collaboration from the domain of computational intelligence, the social sector could make improvements in decision-making processes, data analysis, task optimization, and modeling of complex social problems [2], [3], [4]. The vulnerable population consists of the subgroup that is economically disadvantaged and suffers from chronic physical or mental health problems and a history of alcohol or drug abuse. This group also includes low-income households, youth with abusive families, refugees, and racial or ethnic minorities [5].

Methods of computational intelligence, such as machine learning, are an interesting concept that has gained significant attention in recent years [6]. The advancements in computational technologies have enabled researchers and practitioners to create new systems which aim to improve healthcare systems, government infrastructure, public transportation, information security, and communication networks [7], [8].

Noticeably absent from the subject matter of these studies are social problems, such as homelessness and poverty. Homelessness is a phenomenon caused by various complex circumstances that are outside of the scope of this study; however, economic crises and rapid urbanization are among some of the main factors causing homelessness [9], [10]. Homelessness is a consistent problem throughout all developed nations and has been identified as an urgent global crisis by the United Nations [11]. Recent statistics show that in Canada at least 35,000 individuals are experiencing homelessness every night [12], [13].

Social phenomena, like homelessness, are very complex in nature and thus simulation tools made for social problems are also highly complex. Such applications provide a large number of components and complex features to enable users to create realistic models to reflect their given scenarios. Applications that are made by theoretical computer science researchers are mostly produced for the novelty of the algorithms or comparing the computational resources required to solve large-scale problems. Hence, their user interfaces, if any, are directly related to their program functionality. As a result, the new users' perspectives, abilities, and understanding in relation to software use are often disregarded. These types of applications usually consist of multiple windows for each coded module, simple inputs to obtain user data, buttons to produce results, and additional steps to run the programs. These types of algorithms are usually executed from a code-base and require a prior computer and programming knowledge. Given the required level of knowledge assumed to be needed to use such applications, it can be inferred that the majority of users who engage with these applications are trained individuals, graduate students, and researchers from engineering or computer science fields. Due to these programs being created with a trained user demographic in mind, usability problems arise where barriers for users from different fields become evident. The largest of these usability barriers being the ability to easily navigate and use such applications. The disconnect between existing software and researchers from non-computer science disciplines further emphasizes a disconnect between the computer science community and other fields such as the social sector.

1.2 The Approach

In this thesis, we will touch on multiple sub domains of computer science; namely, Artificial Intelligence (AI) supported web-based software user interfaces (UIs), UI usability testing methods, and computational models that can help mitigate some of the issues that arise in homelessness resource management.

In chapter 3, we present an exploratory study involving three senior professors from multidisciplinary fields. This chapter provides valuable insight for researchers and software developers who are attempting to create a similar graphical user interface (GUI) for social simulations. This chapter provides a detailed description of how the HOMVIZ application was designed, developed, and evaluated. The HOMVIZ user-perceived usability was evaluated with 14 participants using focus groups and questionnaires. The usability testing session provided valuable information about the relationship between computers and human behavior and helped us improve our system. During this usability study, we learned that moderated usability surveys can be time-consuming and difficult to manage. Alternatively to a moderated session, an unmoderated session, which we hypothesized would be easier to conduct, can be used. However, the absence of a moderator might result in unwanted outcomes such as participants' careless responding (CR). Hence we set forth a usability study to compare the data quality obtained from both moderated and unmoderated usability testing sessions.

Chapter 4, showcases our effort to compare remote moderated and remote unmoderated usability testing sessions for the HOMVIZ platform. This survey was approved by the Lakehead University's Research Ethics Board (REB) and was conducted with a randomized control trial with a total of 72 participants and provided valuable information.

Chapter 5 presents a computational model that was meant to showcase the possibility of mitigating homelessness by utilizing computational intelligence. In this chapter, we explore how technology can be harnessed to increase satisfactory and appropriate housing placements. This was done by introducing eight novel heuristic algorithms that create a desirable homeless-to-housing assignment that considers the individuals' characteristics and the nature of the services available. We discuss the efficiency of each of the algorithms through simulations. The algorithms are compared in terms of execution time, solution accuracy, fairness, and the relative difference with the optimal solution of this NP-hard problem.

Chapter 2

Background

2.1 Homelessness

Urbanization is a migration phenomenon in which large populations migrate from rural areas to urban regions. Currently, half of the planet's population is living in urban areas. It is expected that within five decades seven out of every ten humans will be living in urban areas [14]. Rapid urbanization is among the primary factors that cause the rise in the population experiencing homelessness [15]. Additionally, the 2008 economic crisis, which was followed by an economic downturn, is understood to be a primary factor in the recent rise in homelessness [16]. Homelessness has been a persistent and important issue across most developed nations. It has been estimated that at least 235,000 Canadians experience homelessness at some point over the course of a year [17]. Among those affected, there seems to be a disproportionate number of men, young adults who have gone through the youth protection system, Indigenous, and LGBTQ2S+ who experience homelessness. Homelessness, especially street homelessness, is also associated with increased rates of mortality [18]. Homelessness is associated with a number of other adverse outcomes, including substance abuse and justice system involvement [19]. In Canada's larger cities, people experiencing homelessness who also have a mental illness have been estimated to cost governments approximately $60,000 per person, per year in health, social, criminal justice and other services [20]. In total, homelessness has been conservatively estimated to cost the Canadian economy about $7.6 billion annually [21]. The Canadian government has initiated several programs to provide appropriate and safe housing to homeless individuals. Individuals experiencing homelessness may spend one or more nights in a variety of places, including street locations, emergency shelters, violence against women (VAW) shelters, and various forms of transitional housing (transitional housing typically includes on-site support staff and stays are limited to a maximum duration that ranges from a few months to as many as 5 years). Although current policy directions seek to reduce reliance on emergency shelters and transitional housing and quickly transition people who are homeless directly into permanent housing (often with support), emergency shelters and transitional housing remain essential components of the homelessness service system across Canada. On a given night, over 14,000 Canadians are estimated to find themselves in an emergency homelessness shelter. In Montreal, the number of people in transitional housing exceeds those in emergency shelters [17], [19].

2.2 Social Simulation Software

This section will discuss the concept of social simulations and their application in a variety of settings. According to Squazzoni et al. [3], social simulations are often used to combine agent-based computational modeling and knowledge gained from a social science perspective in order to address complex social issues. Through the use of social simulations, researchers hope to develop a deeper understanding of a phenomenon and the ability to investigate the effects of different policy changes and technologies under consideration. A distinct advantage of using social simulations is that researchers are able to further comprehend the degree to which heterogeneity affects a particular phenomenon or issue. Heterogeneity, meaning here that individuals may not all share the same beliefs, preferences, skills, and responses to a social program, plays a significant role in the extent that policies are ultimately effective. When studying and predicting behaviors, heterogeneity often presents the largest challenge and its effects have been largely overlooked in past research [22].

While it is important to consider the use of social simulations in training in a variety of disciplines, research conducted by Mago et al. [4] developed simulation methodology that targets homelessness. Their research is intended to demonstrate that through the combination of computational modeling and social simulations, various strategies for allocating resources to address homelessness can be tested. As such, this section focuses on the related work dealing with social simulation targeted at homelessness. Mago et al. describe homelessness as a state of being without permanent housing but are careful to consider that homelessness can be experienced in several different ways. For example, one can be living in emergency shelters, temporarily with acquaintances or friends, on the street, or in housing that is considered to be unfit for habitation [23], [24]. Given the heterogeneity of the homeless population, and of the causes of homelessness, it is difficult to develop a precise solution or approach for dealing with the issue. Combining health and social science research findings on the effectiveness of programs with methods from a computational background, complex social issues such as homelessness may be addressed more effectively. Their research investigate methods using Fuzzy Logic and Fuzzy Cognitive Mapping to identify some of the most important causes of homelessness. Fuzzy Logic is an approximate technique that allows variables to have a truth-value between 0 and 1. This method enables researchers to assign a numerical weight to what might otherwise be considered subjective value. Once a numerical value has been assigned, the values can be mapped on a graph to create the Fuzzy Cognitive Map (FCM) which will demonstrate the nature of the relationship between the two connected nodes (or variables) of a graph. This method of investigating the relationship between several variables is particularly useful when studying issues relating to homelessness and social sciences in general because it can empirically graph complex data in a way that is dynamic and adaptable. Additionally, they suggest that the FCM can help researchers and policymakers identify variables that may have more significant impacts in a complex system such as affordable housing, social support services, family support services, addiction, and mental illness supports. If these issues can be identified, then a more strategic policy can be developed to help approach them.

Fowler et al. [11] investigate the various approaches and challenges of homelessness screening and resource allocation. For example, the Vulnerability Index— Service Prioritization Decision Assistance Tool (VI SPDAT) is a tool that is used to categorize those seeking homelessness assistance based on vulnerability levels. However, they suggest that the validity of this tool has not been demonstrated and therefore fails to assess needs in an accurate manner and ensure that services match the needs of screened individuals. This study also mentions the difference between the Housing First initiatives and Treatment First services. Housing First initiatives provide users with a combination of access to housing and additional support at the first opportunity. Treatment First services provide support to those seeking assistance, however, individuals are expected to manage the underlying conditions contributing to their homelessness before they are considered ready and able to manage to live in their own housing [11], [25]. Due to the extremely complex nature of homelessness, Fowler et al. suggest that a complex system may "offer a unique tool for evaluating coordinated responses" to homelessness and associated resource allocation. In relation to computer simulations, they assess assumptions of existing systems and can help identify additional areas in homelessness relief initiatives where beneficial interventions can be made.

2.3 User Interface Design

This section highlights the methods that have been used in past research to ensure that web designs are highly functional for users. In order to create functional and desirable web designs, designers need to have a good understanding of how users interact with different web platforms and also consider the impacts of using different types of devices. Page et al. [26] have contributed to this conversation by investigating the "complexities involved in measuring usage and interaction with the web" through the use of self-report surveys. Self-surveys are valuable to web designers because they can help them further understand the relationship between consumers and their perception of different types of software. The technology acceptance model (TAM) has been very popular for explaining the correlations between user perspective and the use of technology. TAM consists of two categories, "perceived ease of use" (PEOU) and "perceived usefulness" (PU). It is believed that there is often a strong correlation between how easy a system is to use, how useful that system is, and the frequency of the system's use by the consumers. Therefore, if a website is perceived to be useful, this can increase the likelihood that the website will see frequent use.

Another method for determining how to produce a well-designed web platform is through the Unified Theory of Acceptance and Use of Technology (UTAUT) which was developed through several adaptations of TAM. Al-Qeisi et al. [27] propose that website design is multidimensional and has a hierarchy of needs in terms of which elements are the most important for the functionality of the design. They argue that when using the UTAUT model to conceptualize web design and functionality ideas, the quality of those designs "outperform existing models." Additionally, they feel that improving a system's appearance can often increase an evaluation score thus resulting in more overall usage. The UTAUT consists of four components: performance expectancy, effort expectancy, social influence, and facilitating conditions. "Performance expectancy" refers to how a particular technology helps a user completes a particular task while "effort expectancy" refers to the level of difficulty the user is expected to manage while using the technology. "Social influence" refers to the recommendation of the technology from the user to others around them and the level of belief that this technology should be adopted by others. Finally, "facilitating conditions" refers to the user's understanding and perception of the support resources available to them. Al-Qeisi et al. found that perception of the quality of website design can influence the user's behavior. Their results are consistent with previous research which found that design features including access, navigation, and speed are determinants of software "adoption and perceived usefulness" [27].

An additional method for providing information about user interaction with a particular system is through implementing mouse tracking heatmaps. Heatmaps provide key insight to designers when constructing a website as they allow designers to better understand where users are spending most of their time when navigating a particular software. Katsano et al. [28] explore information scents and their impact on the user's behavior while navigating a website. The term "information scent" refers to the hint that a link title provides to the user regarding the functionality of that link. Specifically, they are interested in how they impact the "distribution of attention, confidence in the choice of link, efficiency, and effectiveness" of the user. They presented heatmaps of their participant's behavior while navigating various websites in order to determine where inefficiencies existed in the designs.

Testing GUI is vital when designing software due to the fact that GUIs are often most commonly used in consumer software. Designing GUIs that can accommodate a number of different "event sequences" can be difficult [29]. A rather simple, yet perhaps underestimated, tool for producing functional and desirable GUIs is by having competent testers. Chen et al. [29] study the potential methods of improving GUI testers and throughout their research propose solutions such as making sure testers are aware of various strategies and best practices while testing, providing testers with "in situ guidance" to reduce the amount of decision making while testing, and ensuring that those who are hired to test GUI are adequately experienced.

2.4 User Interface Evaluation

This section will discuss the methods used in similar research projects when evaluating UIs. There are a number of ways that UI can be evaluated to improve system usability. The most popular of these methods include surveys, interviews, and focus groups.

Structured tools are required when evaluating UI functionality in order to gather data that is as useful as possible. Many different usability questionnaires exist for this function. Among the most popular are the System Usability Scale (SUS), the Post Study System Usability Questionnaire (PSSUQ), the Computer System Usability Questionnaire (CSUQ), and the After-Scenario Questionnaire (ASQ). The System Usability Scale, or the SUS, is one of the most commonly used questionnaires [30]. The SUS is a 10-item questionnaire that uses a 5-point Likert scale ranging from 1, strongly disagree, to 5, strongly agree. To calculate the SUS score, the raw score of even-numbered items must be subtracted from 5 and 1 must be subtracted from the raw score of the odd-numbered items. The sum of these scores must then be calculated and the result is multiplied by 2.5 to reach a "standardized SUS score". A standardized SUS scored above 85 indicates a "highly usable" system, a score between 70-85 "is characterized from good to excellent," and a score from 50-70 indicates "that the system is acceptable but it has some usability problems and needs improvement," and a score below 50 indicates that the system is "unusable and unacceptable" [31].

Coppers et al. [32] explore the evaluation process for a UI that aids professional translators in their work. They had 13 participants test one version of their UI and another 13 test a second version. Through the use of a remote usability tool, they recorded the participants' screen time during their test and followed that with a usability questionnaire. In this instance, the SUS questionnaire was used. In the first version of their UI, the SUS score was 60.0 and in the second version, it was 67.5, ranking their scores as, respectively, fair and good.

Marky et al. [33] studied the usability of the Benaloh Challenge, a popular method of "cast-as-intended variability" when voting in elections. To investigate the usability of the Benaloh Challenge, they used the SUS questionnaire as well as included two open-ended questions. Two versions of the UI exist, an automatic and manual version. The automatic version received a SUS score of 79.4 and the manual version received a 75.4, ranking them both under the SUS scale as good. Considering the existing literature regarding usability surveys, for the purposes of our research, it was determined that the SUS questionnaire would provide the best data for a full analysis of our system.

However, some choose not to use a usability questionnaire to evaluate their UI. For example, Hassib et al. [34] published a study that investigated the proponents of implementing heart rate information into chat applications. In order to collect data regarding the usability and implementation of heart rate monitors when using chat applications, they conducted a focus group with 14 participants. Their focus group consisted of two parts. First, they asked users how they would normally convey their emotions while using a chat app and what they felt the inadequacies of the existing methods were. Second, they introduced the concept of using heart rate monitors to convey emotions while using chat apps and asked the users to brainstorm ideas of how to implement a GUI with this concept. The researchers conducted a thematic analysis of the seven hours of recorded interview data.

Hudson et al. [35] investigated the barriers that children face when using non-academic related 3D design software. Eight facilitators with experience working with children helped conduct semi-structured interviews with 15 participants who completed tasks onto the Tinkercad 3D design application. Each interview was recorded, transcribed, and then studied for common themes. They used an inductive analysis approach to study the interview transcripts, open coding to label the transcripts, and affinity mapping to identify common themes. Additionally, screen and audio recordings from the child's interactions with the software were analyzed to identify instances where more thought or problem solving may have been required.

Chapter 3

Design, Development and Usability Evaluation of a Web-based AI-supported Simulation and Modeling Software for Homelessness

All of this chapter is submitted to a peer-reviewed journal as:

- Khoshnevis, P., Latimer, E., Tillberg, S., Aubry, T., & Mago, V. (2020). Design, Development and Usability Evaluation of a Web-based AI-supported Simulation And Modeling Software For Homelessness.

Over the course of two years and countless hours of collaboration with domain experts from multidisciplinary fields, did this work come to compilation. The design, development, and evaluation of our software is explained in detail to provide valuable insight for researchers and developers with the same interests

Keywords: User interface design; User interface evaluation; Usability testing; Software interface;

3.1 Introduction

Over the past three decades, homelessness has become an increasingly problematic phenomenon in Western industrialized countries and is often associated with economic and social marginalization [36], [37]. Multiple societal factors have been identified to be contributing to the problem including the growing economic gap that is resulting in a larger proportion of the population living in poverty, a shrinking pool of affordable housing in cities, and ineffective social programs and policies targeting homelessness [38], [39], [40]. To address this problem, various government programs have been developed to assist homeless individuals to access housing, healthcare, and social services [41], [39], [42]. Given the size of the homeless population worldwide and a growing research base on the effectiveness of programs targeting homelessness, policymakers are faced with determining how best to expend resources for these programs in the most efficient manner. Simulation and modeling software are examples of these alternative methods that can help future planning and optimizing available resources. A simulation and modeling is a means of imitating a real-world environment to discover the origin of known or unknown events [3] or in some cases, it could be a means of identifying the common properties of similar or related events [43]. A computer simulation consists of algorithms and mathematical expressions that represent the behavior of a system in a real-world environment and is beneficial in terms of efficiency and scalability [44], [45]. Conducting simulations on a large scale, produces an abstract overview of events that can be used to understand and map the macro-micro level phenomena [3]. These events can represent any real-world phenomenon ranging from economic to social events. Simulation software that is built to uncover a social phenomenon is referred to as *social simulation software* [46].

In this chapter, we present a GUI, namely the Homelessness Visualization (HOMVIZ) platform. The HOMVIZ platform is a responsive web-based application that enables researchers or policymakers to easily create simulation models that predict future trends in homelessness and obtain resource utilization analysis. Our research goal evolved while designing a UI to represent our deep learning algorithm. Given the complexity of homelessness, our simulation model can have up to 1.2m people as the model's population, four types of resources of homelessness management, and three potential living situations with many dynamic properties within each. Due to the large number of features and properties, we soon realized designing an easy to use UI is a challenging task. Therefore, we started our design process aiming at answering the following: (1) is it possible to hide the complexities of an AI algorithm for a social system using a simple web-based interface?; (2) is it possible to build a simple to use application that captures the intricacies of a complex social system like homelessness?

Our contribution is four-fold: (1) a rich UI that hides the complexity of the underlying algorithm and simplifies the process of creating social simulation models; (2) a detailed explanation of the design and development process; (3) a combined analysis of the quantitative and qualitative usability survey data; and, (4) freely available source code of the HOMVIZ platfom[1].

3.2 Related Work

In the related work section we will discuss methods that were implemented by researchers to evaluate and design effective UIs. Hong et al. [47] speak to the importance of "systematized evaluation criteria with a strong theoretical basis" in their research. To achieve this, they propose the development and implementation of evaluation criteria that are based on

[1]https://github.com/pedramvdl31/HOMVIZ

the same evaluation approaches of architects designing buildings. They suggest six criteria are necessary for this: internal reliability, external security for structural robustness, useful content, usable navigation for functional utility, system interface, and a communication interface for aesthetic appeal. Through their comparison of web design and building construction, they explain that both entities provide users with a space to perform specific activities. For example, where one provides a physical space such as a classroom and the other provides virtual space such as a virtual classroom. Therefore, both physical buildings and websites require an architect to design and create functional spaces that are structurally robust as well as aesthetically pleasing. Following a survey, they found that the six criteria had different impacts on user satisfaction for different websites and these impacts were classified into two categories: user goals and user activity levels. They concluded that designing a test website using the six criteria produced better user experiences.

As internet technologies continue to advance, the need for evaluation tools which prioritize the creation of user-centered platforms has also increased. Since the 1980s researchers have published a number of evaluation measures that are specifically targeted at discerning the usability of web-based platforms and applications [30]. Some notable usability-testing questionnaires are STRATUS, Post Studies System Usability Questionnaire (PSSUQ), Computer Systems Usability Questionnaire (CSUQ), Technology Acceptance Model (TAM), and System Usability Scale (SUS). STRATUS is a questionnaire that focuses on strategic usability [48]. It is separated into two parts, the first of which involves the collection of participant's demographic information and the second part focuses on the assessment of strategic usability. The questions are a combination of closed-ended questions ranked using a five-point Likert scale, multiple-choice, and some open-ended questions. The second part of the questionnaire is divided into five blocks in which the types of questions used are more clearly organized [48]. The popularity of STRATUS has been correlated to its cost-effectiveness, minimal resource use, and time efficiency [48].

The PSSUQ and CSUQ are closely related as the CSUQ is based on the PSSUQ. Initially, the PSSUQ was an 18-item questionnaire but was later revised to create the 16-item CSUQ. The main difference between the two questionnaires is that the CSUQ is believed to be broader in scope due to its wording whereas the PSSUQ is more closely associated with direct testing environments. Due to its broader reach, the CUSQ lends itself better to mail-out survey scenarios and as a result allows for a larger sample size [49]. Both the PSSUQ and CSUQ use a 7-point rating scale varying from strongly agree to strongly disagree with an outlier option of “not-applicable” [49].

Another popular usability rating system is the TAM. The purpose of TAM is to investigate the user’s “perceived ease of use” and “perceived usefulness” of a particular system [50]. This has been a beneficial model when researchers are specifically exploring what effort is required from users when integrating certain technologies into their regiments [50].

Finally, there is the SUS. While SUS was one of the last proposed usability system surveys to be published, it has become one of the most popular [30]. The SUS questionnaire is primarily used to investigate levels of user satisfaction and learnability for web applications [50] and has proven to be extremely useful in “task-based usability studies” [30]. The SUS is a 10-item questionnaire in which the questions are divided into five positive statements and 5 negative statements [31]. While there are some debate and variance between different researchers on the test’s coefficient alpha, which is the numerical value that determines the accuracy of a usability test, the SUS proves to be an accurate measure of a system’s perceived usability [30].

3.3 Methodology

HOMVIZ Platform Overview

The HOMVIZ platform is a GUI that enables users, namely researchers and policymakers, to easily use a deep learning algorithm, modified deep q-learning (MDQL), and modified neural fitted q-iteration (MNFQ) which has been more succinctly named: the BEAUT model [51]. The BEAUT algorithm is originally based on the Markov decision process. The Markov decision process uses transitional probability matrices to predict the movement of individuals between states. However, for complex simulations with a large number of parameters, calculating an accurate transition probability for each moving cycle (weekly in our case) is very costly. The BEAUT algorithm uses deep learning to dynamically generate a transitional probability that best reflects the data in hand. The BEAUT algorithm uses the data derived from the Montreal site of the *At Home/Chez Soi* project [41], [52]. This chapter only covers the design, development, and evaluation of the HOMVIZ application and does not discuss the algorithmic approaches of the BEAUT algorithm.

In addition to a standard registration and login page, the HOMVIZ platform consists of three pages named the *simulation homepage*, the *create simulation* page, and the *simulation results* page. The *simulation homepage* provides the user with a list of previously created simulations, additional controls to delete or view simulations, and useful navigation links. The *create simulation* page provides the main functionality of the HOMVIZ application. On this page, the user can create a new simulation model using the simulation stepper wizard which consists of 5 steps, shown in Figure 3.1, 3.2, 3.3, 3.4, and 3.5. The stepper wizard was added to the HOMVIZ application to enhance the user experience by simplifying the process of creating a new simulation model. Each step validates the user's inputs and provides feedback if the input pattern does not meet the expectations. Every step provides a *Next* and a *Back* button. By default, the *Next* button is disabled until the user inputs all the required values or parameters. On every step, we provide additional information about parameters and inputs using interactive tooltips [53], information dialogue windows [54], and tutorial videos using a slider window.

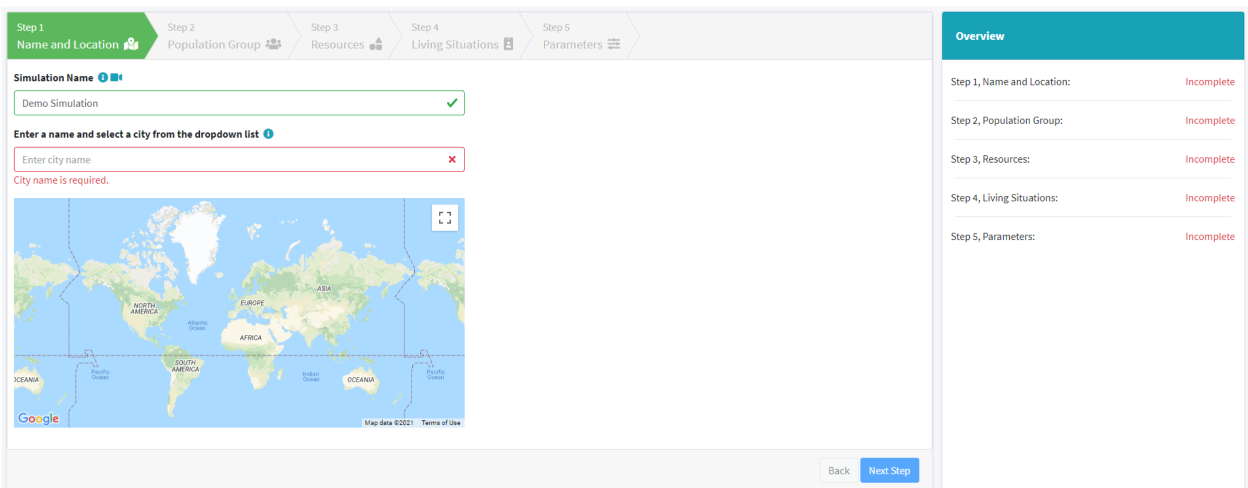

Figure 3.1: On the first step, the user is required to provide a simulation name and the geographic location of their population of interest.

On step 1 (Figure 3.1), the user must input the simulation name and the name of the city for which they wish to create a simulation model. The simulation name must be a string with a length ranging between 3 to 20 characters. To enhance the user experience the city name input uses the Google map API's auto-complete function that provides complete city names as the user is typing. Once a city is selected it is shown on a small world map window below the input field. This is done to improve the aesthetic appeal of the platform.

On the second step (Figure 3.2), the user must select at least one population type that is provided to them using a drop-down list. The population types are used to simulate the movement of individuals (population) between resources and living situations. Each

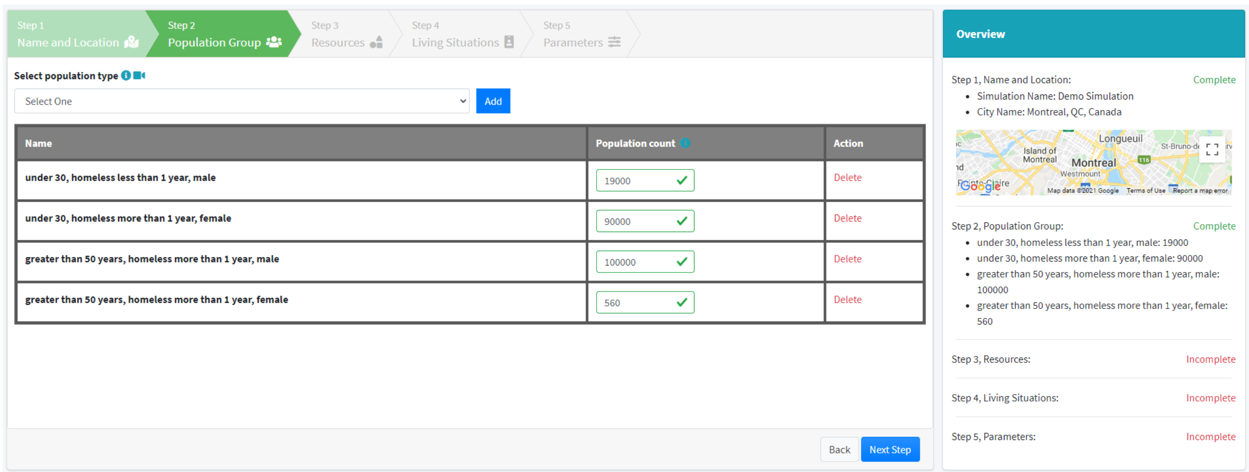

Figure 3.2: On the second step, the user is required to provide population types and counts.

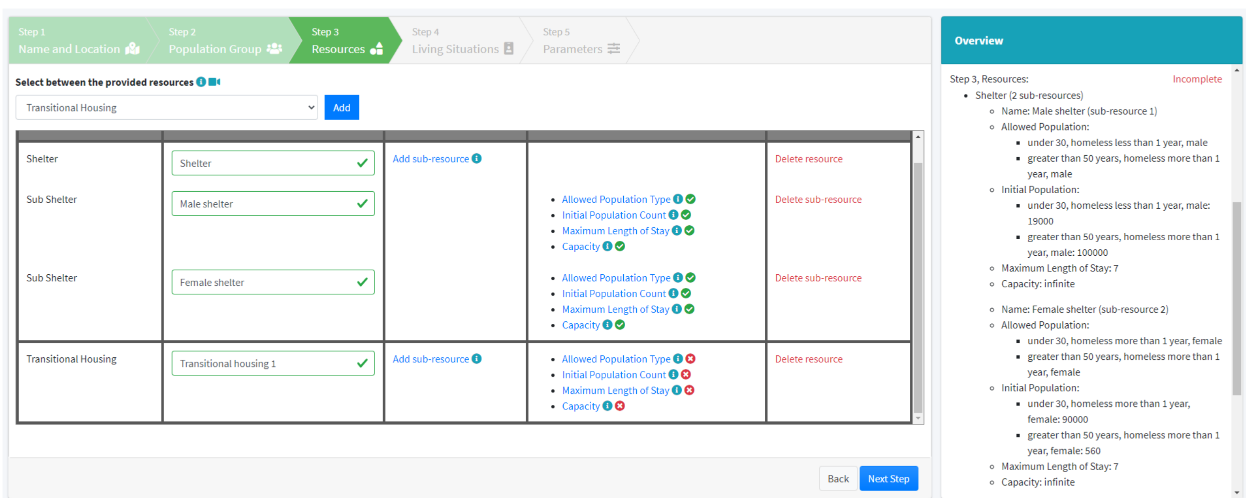

Figure 3.3: On the third step, the user is required to provide the allocated resources and metadata for their simulation model.

population type can be selected only once and can be deleted at any time. Once added, the table located below the drop-down will be appended with that population type. The reason for the user's inability to add custom population types is that the BEAUT algorithm was trained on a dataset with specific population types. Once the user has added a population

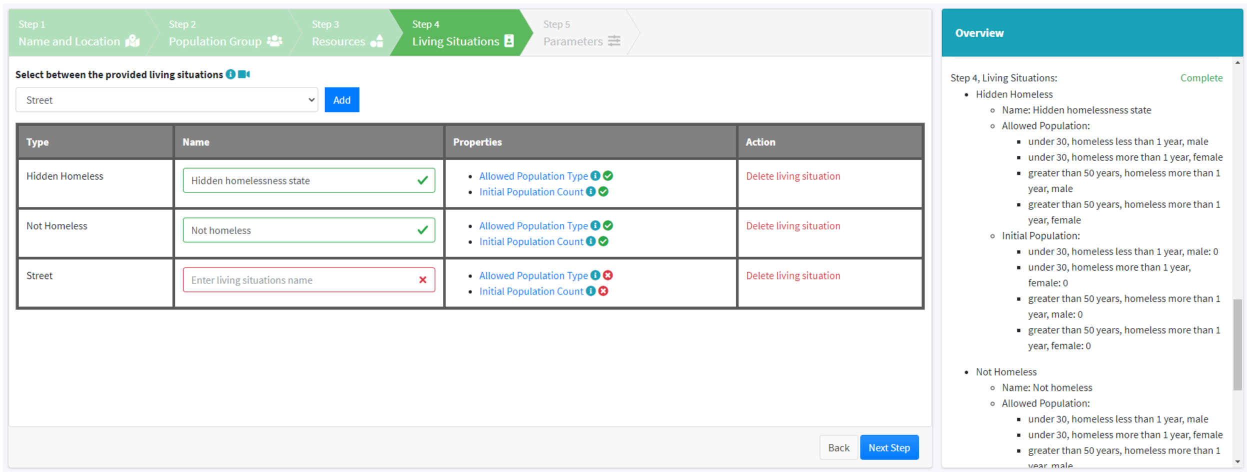

Figure 3.4: On the fourth step, the user is required to provide the living situations and metadata for their simulation model.

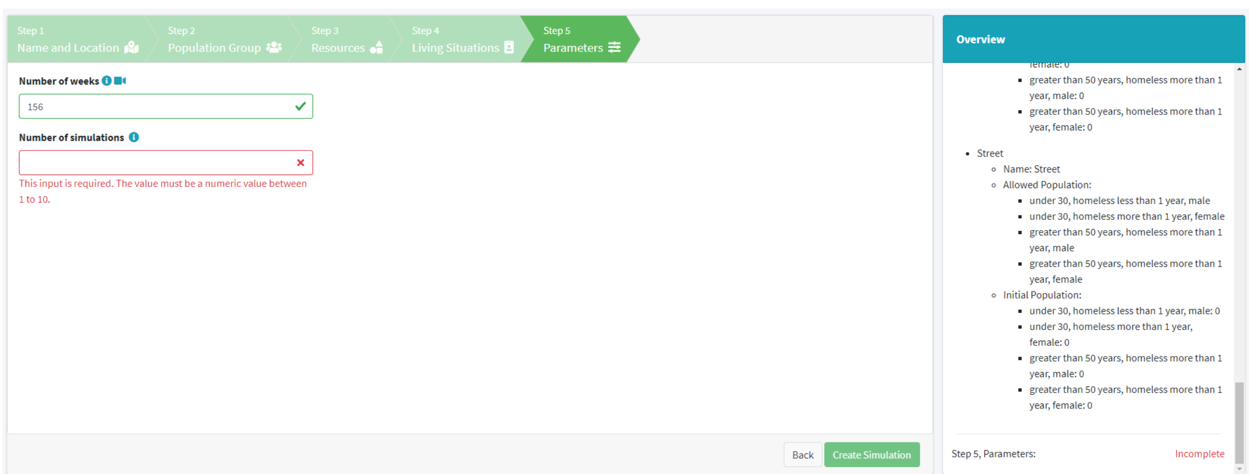

Figure 3.5: On the fifth step, the user is required to provide the duration of the simulation, and the number of simulation runs.

type they must also enter the population count for that type. The population count is required and expected to be a numerical value ranging between 1 to 100,000.

On the third step (Figure 3.3), the user must select at least one resource from the drop-down list provided to them. A resource is defined as a type of facility where people experiencing homelessness stay safely overnight. The included resources for this simulation algorithm are *Hospital*, *Shelter*, *Transitional Housing*, and *Addiction / Rehabilitation Center*. Users can select any of the provided resources and add cohorts to each resource if necessary. In the HOMVIZ application, we used the term sub-element to refer to the cohort. For example, a shelter resource can be further divided into two sub-resources where one might be a male-only shelter and another a female-only shelter. Once added, a resource or a sub-resource has multiple properties that must be filled before moving to the next step. We added interactive tooltips that are either green with ✓ icon or red with ✗ icon. If the user hovers their mouse cursor over the interactive tooltip, a message will appear indicating whether the property is set correctly or not. Using these visual indicators, users are reminded which element needs their attention. These properties are *Allowed population type*, *Initial population count*, *Maximum length of stay*, and *Capacity*.

Even though the scope of this study is the user interface of the HOMVIZ platform, we decided to provide a brief explanation about different resources and parameters to provide a better understanding of the problem that the system is attempting to solve. The following is a brief explanation of the resources that the user can add in step 3.

- **Hospital:** Hospital is a public or private entity that provides health care to homeless individuals (among others of course).

- **Homeless shelter:** Homeless shelter is a homeless service agency that provides temporary basic overnight accommodation to homeless individuals. This includes shelters for women who are victims of family violence.

- **Transitional housing:** Transitional housing is a residential setting in which individuals may live, usually with on-site supports, for a limited amount of time (which can be as long as 5 years).

- **Addiction/rehabilitation centers:** Addiction/rehabilitation centers are a type of facility in which those with substance use problems live and receive treatment for a specified period of time.

The following is a brief explanation of the parameters that must be added for each resource and living situation in steps 3 and 4.

- **Allowed population type:** This property mandates which population type (selected by the user in step 2, e.g., men, women) can enter this resource.
- **Initial population count:** This property is the total population count that is residing in this resource at the beginning of the simulation.
- **The maximum length of stay:** The maximum length of stay is the maximum number of weeks that one individual can continuously stay in this resource.
- **Capacity:** Capacity refers to the maximum number of individuals that can reside at this resource at the same time.

On the fourth step (Figure 3.4), the user is given the option to select from three different living situations. The simulation algorithm is designed to include *Hidden homeless*, *No longer homeless*, and *Street* as the defined homeless living situations. Similar to resources, living situations, when added, have multiple properties. These properties are *Allowed population type*, and *Initial population count*. The following is a brief explanation of the provided living situations:

- **Hidden homeless:** This situation refers to homeless individuals without guaranteed accommodation who temporarily live with relatives, friends, or acquaintances. These individuals may not appear to be homeless.

- **Not homeless:** This situation refers to when a homeless individual moves out of the state of homelessness and has their own place to live for as long as he or she wishes.

- **Street:** This situation refers to homeless individuals living on the street and who are usually identifiably homeless.

On the fifth step (Figure 3.5), the user must fill in two inputs: the *Number of Weeks*, and the *Number of Simulations*. The *Number of Weeks* is the length of the simulation and the *Number of Simulations* is the distinct number of simulation runs. The overall accuracy of the simulation will increase as the number of simulation runs increases. The deep learning algorithm uses the given information and predicts the number of homeless individuals moving between different resources and living situations in a weekly cycle. Following the completion of their simulation build, the user will be redirected to the homepage where a progress bar and interactive notification badge will provide real-time updates on the algorithm's progress processing the simulation model.

On the *simulation results* page, the user can view the simulation result using the provided interactive graphs. The results are presented to the user using line graphs, web graphs, and bar graphs using two different metrics. First, the population counts in resources and living situations over the given period in a weekly cycle (Figure 3.6). Second, the initial population counts (population counts at the beginning of the simulation) in comparison to the final population counts in each resource and living situation (Figure 3.7, 3.8). The movement of population across the model is shown for each population type in separate graphs and for all population types combined. The graphs are cross-browser compatible and highly interactive. The user can view the information in detail by hovering over a specific location on the graph or add or remove data groups to compare specific data points.

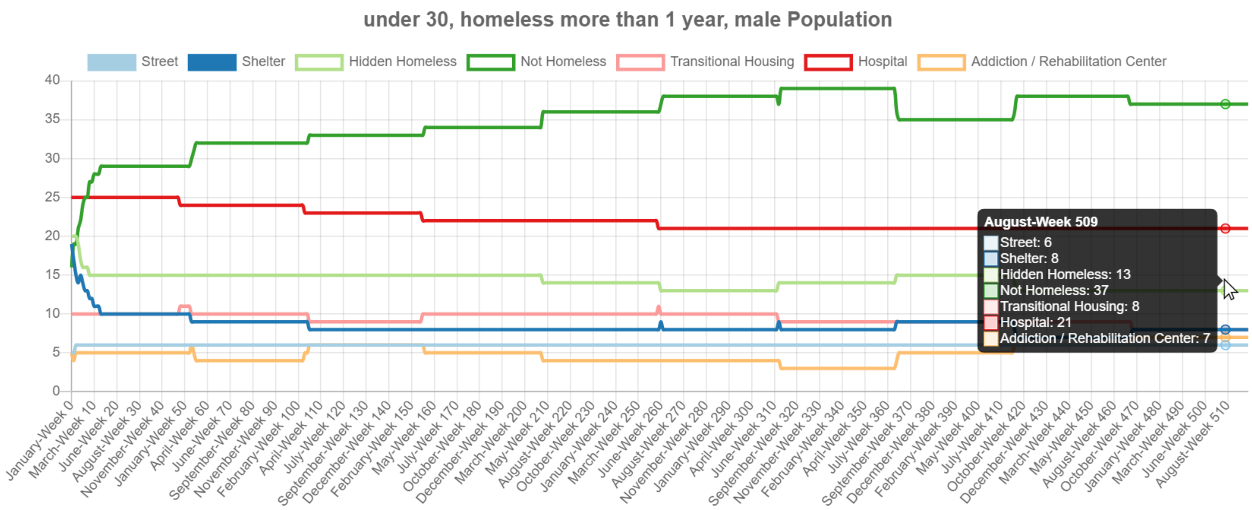

Figure 3.6: Graphical representation of simulation result using a line graph.

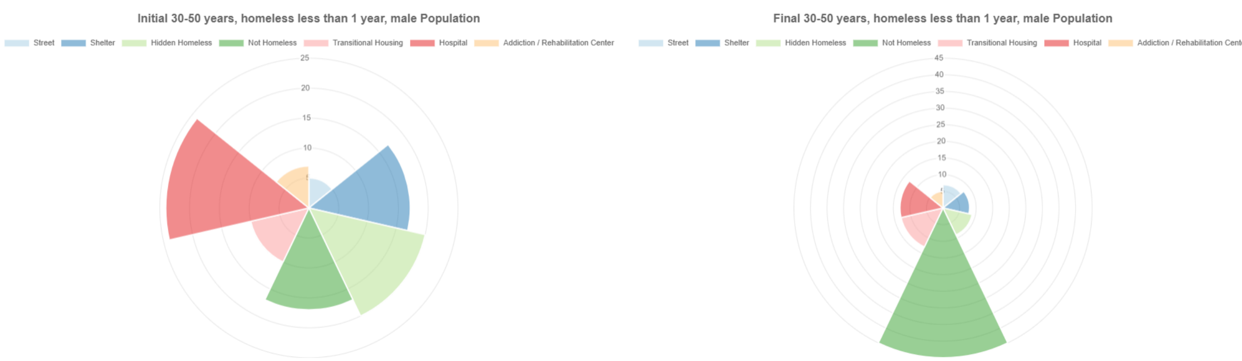

Figure 3.7: Graphical representation of simulation result using a radar/web graph.

System Overview

The HOMVIZ platform consists of a web server and a browser application developed using the Laravel PHP framework [55] and MySQL relational database for data storage. To ensure security, the server uses standard security protocols that have been provided by the Laravel framework. The user authentication controller is implemented using the client-side and server-side session-based method. The passwords in the database are secured

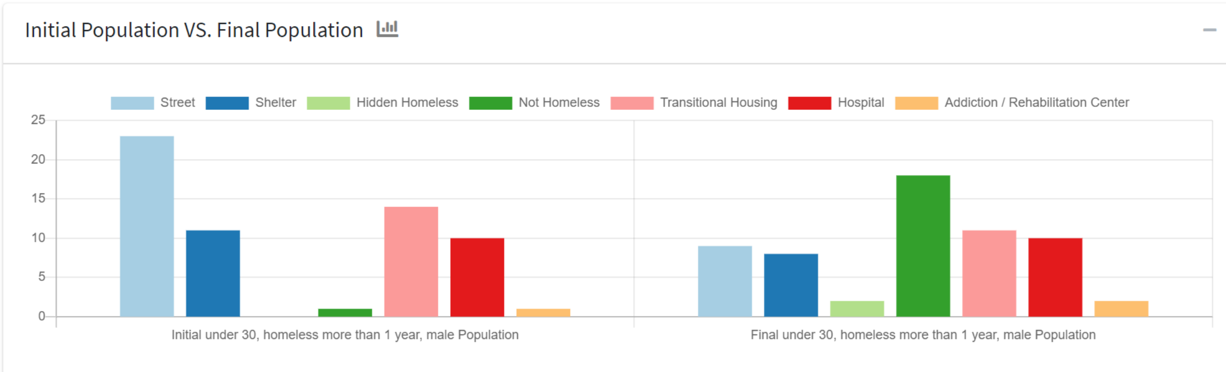

Figure 3.8: Graphical representation of simulation result using grouped bar graph.

using a one-way cryptographic hash function. Additional security measures that have been implemented include cross-site scripting (XSS) protection, forcing HTTPs when exchanging sensitive data, and cross-site request forgery (CSRF) attack prevention. The client-side was developed as a responsive web application using HTML5, CSS3, and JavaScript. The UI styling of the platform was improved and standardized using the Bootstrap framework and toolset [56], and jQuery UI [57]. Visualization graphs were developed using a modified D3.js library [58].

Design and Development

Our original motivation was to create a GUI where users, typically researchers and policymakers, could utilize a deep learning algorithm to create a simulation model to predict trends in homelessness in their community of interest. The initial HOMVIZ simulation prototype was designed using a vertical HTML form. During our testing sessions, we realized that creating a model using our prototype resulted in too many nested elements and a long page that required excessive scrolling.

Due to the lack of direction in our previous application, it was difficult for the user to know where to start, what needed their attention, how to progress through the simulation, and what was expected of them to successfully create a simulation model. Our prototype was a standard UI representing an algorithm solely made to enable the user to interact with our algorithm. We soon realized users who are not familiar with our algorithm or not trained to use our application would struggle using or even understanding our UI. A screen-shot of our prototype is shown in Figure 3.9. To solve this problem, we incorporated a stepper wizard that breaks the modeling process into separate steps, guides users by providing information about most of the elements and validating user input data. To realize our design goals, we relied on a user-centered approach where the end-users are part of the design and development process [59], [60], [61]. Initially, we started the design process using a team of multidisciplinary academic scholars as our users.

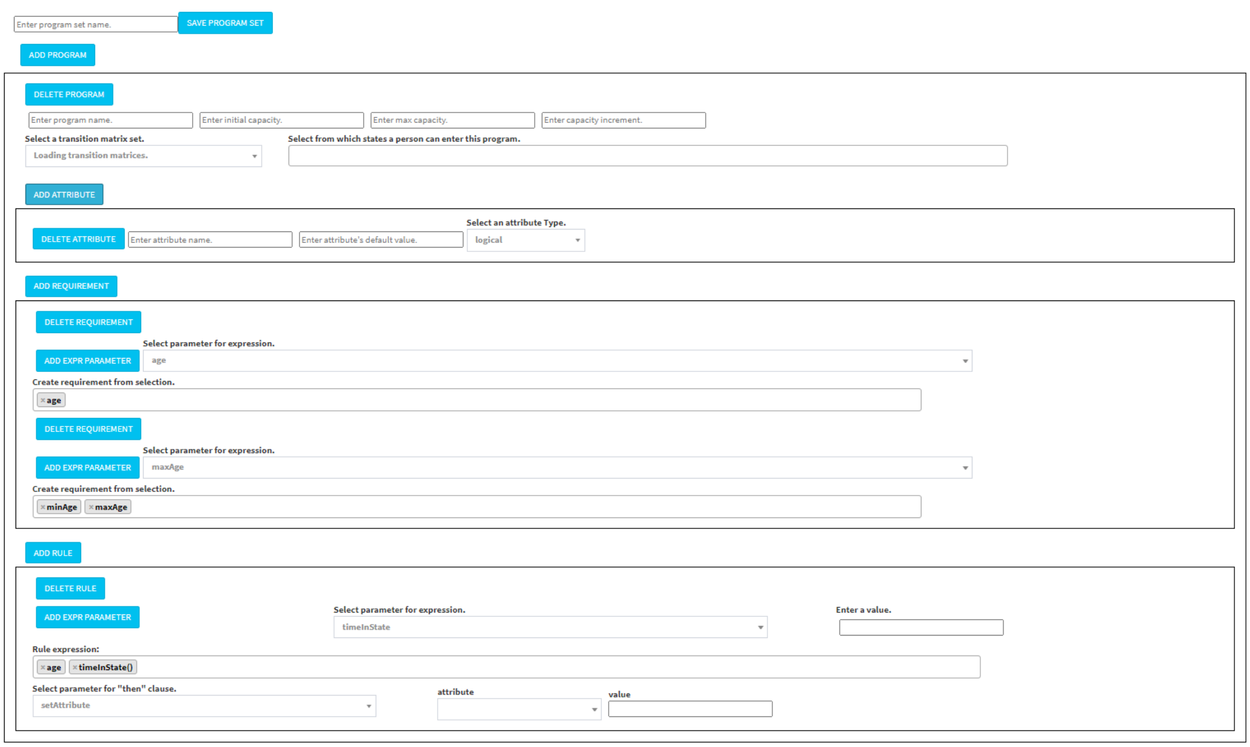

Figure 3.9: A screen-shot of the HOMVIZ initial prototype.

The design and development process of the HOMVIZ platform was completed in four stages. Table 3.1 illustrates these phases of our design and development process.

Table 3.1: Description of the stages of the HOMVIZ platform design and development

Stages	Technique	Purpose	Stage of the Design Cycle
1	A sequence of brainstorming sessions	To design, improve and evaluate our prototype	At the beginning of the design project
2	Weekly software review sessions	To review the software development process and course correction	Early and mid-point in the design cycle
3	Walk-through and usability sessions	Finding design flaws and anticipating alternative design elements	Early and mid-point in the design cycle
4	Usability session, focusgroup, and questionnaire survey	Gathering user feedback in the form of quantitative and qualitative data	Final-point in the design cycle

During *stage 1*, we conducted a sequence of brainstorming sessions where after running through various sample scenarios using paper sheets we drew the prototype wireframes. During *stage 2* and *stage 3*, we gradually developed and improved the digital version of our design. Over the course of two years during our weekly group sessions, we tested the newly developed application by creating sample scenarios and took notes of our participants' comments. The academic members from the computer science department mostly commented on the UI and functionality of the application and members from psychology and epidemiology commented on the technical language and modeling aspect of the application. During *stage 4*, we conducted a user study with a larger participant sample size. Our user study consisted of usability sessions, focus groups, and questionnaire surveys which provided us with more insight into any shortcoming of our software.

To create a user-friendly interface we followed the guidelines provided in research work by Hong and Kim [47] as well as, Thielsch and Hirschfeld [61]. Based on their research, we focused our attention on three main aspects of the system: robustness, functional utility, and aesthetic appeal, as described below:

A robust application can provide internal reliability, credibility, and system security to the user. Internal reliability denotes the operational stability provided to the user ensuring a high-quality user experience. The system security includes the measures taken to provide a secure platform to the user which ensures data integrity and security. A robust application is a product of a well-thought design that is created to handle any expected or unexpected threat or behavior [47]. Our system ensures that the application is internally reliable by using input validation, input sanitization, and a controlled/validated multi-step stepper wizard. The user-input data must meet certain criteria in order to be processed on our system, namely on the platform's client-side, server-side, and on the algorithm's standalone server. For the client-side input validation, we used the standard HTML and JavaScript form constraint validation. The input validations include disabled, min, max, regular expression pattern, required, and type matching. The data were re-checked on the server-side for integrity.

Functional utility refers to the inclusion of appropriate features and functions that are relevant to the user's needs. These features assist potential users in completing their intended activity by providing them with useful information ensuring minimal errors and frustration while using the system. The functional utility can be divided into two items: useful content and usable navigation which are among the most important factors for user satisfaction [47], [26]. Useful content refers to the information quality and informativeness of a platform and it is measured by the relevancy of the information provided to the user to aid them in accomplishing their task [47], [62]. Usable navigation allows the user to easily navigate through the steps or pages provided on a web platform.

According to Hong and Kim [47], the navigation usability is "the ease of navigation of websites." It has been shown, that an application with seamless navigation promotes the feeling of ease of use and puts the user in the state of flow which results in a greater desire to use the platform and helps them to easily accomplish their task [47], [28], [26], [63]. To provide informativeness and clarity to our platform, we used interactive tooltips, information dialogue windows, and video tutorials using a slider window to the right of the screen in our application. The video tutorial provides clear and easy to follow instructions for each step of the simulation stepper wizard.

These methods are used to deliver important information to the user regarding potential errors and additional informative tips to aid them while creating their simulation. The information tips are marked using an icon and the video tutorials are marked using a icon. Figure 3.10 and 3.11 illustrate sample information dialogue windows/modals. Figure 3.12 shows the video tutorial slider window. To promote usable navigation we have placed buttons and links with a similar type of functionality in similar locations and used a uniform color scheme throughout our application which helps the user in finding the desired action button and links. Finally, a stepper wizard was incorporated to provide the user with step-by-step information and navigation.

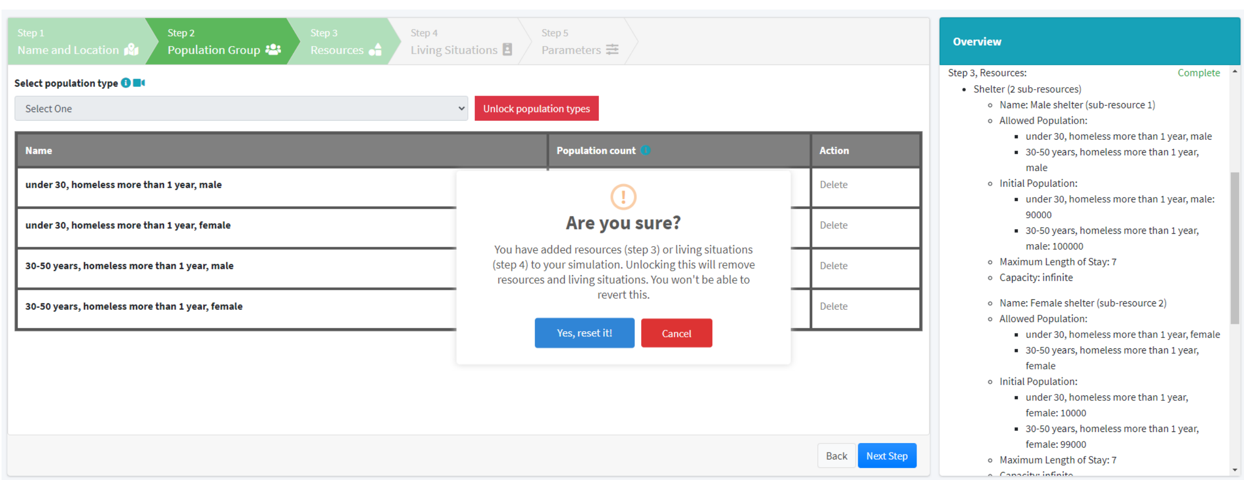

Figure 3.10: A notification dialogue window to confirm and validate user request using confirm and cancel buttons.

Aesthetic appeal includes using features such as graphs, images, colour schemes, and other styling elements that make using the software more enjoyable as well as simple to

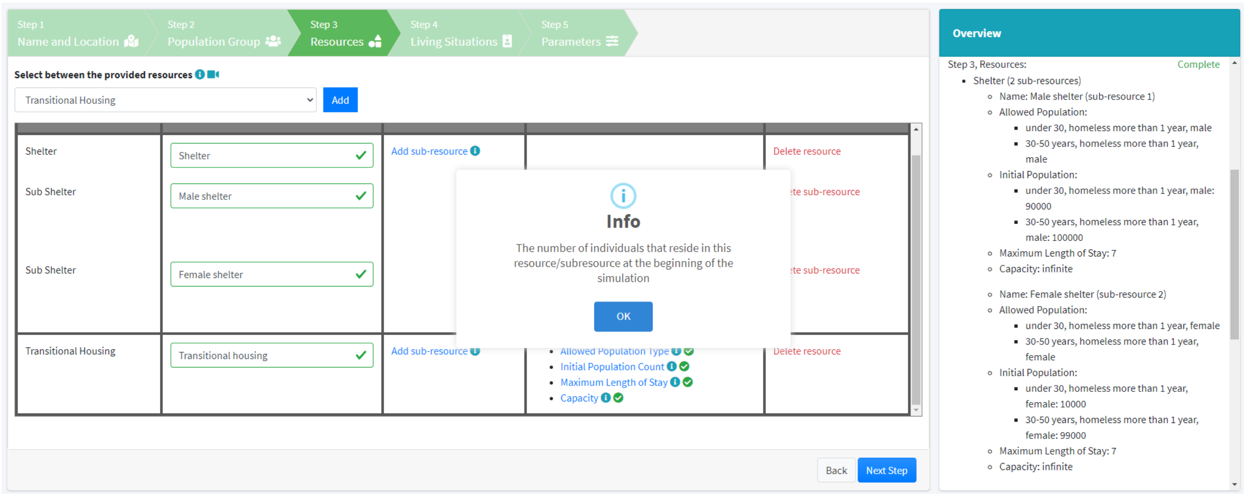

Figure 3.11: A sample information dialogue window that provides additional help to the user regarding the definition of an element; in this case, the initial population count.

Table 3.2: HOMVIZ colour theme

Types	Colours
Positive	Green
Primary	Blue
Negative	Red
Secondary	Gray
Information	Light blue

understand [47]. To create an attractive UI, we used styling libraries such as Bootstrap [56], jQuery UI [57], Chart.js [64], Plotly.js [65] and Sweetalert [66]. Our web platform is responsive, which dynamically adjusts to the user's screen size and is known to improve user experience and promote the feeling of ease of use [67]. To ensure a uniform colour style across our application we used the colour scheme shown in Table 3.2 which was adapted by following the best practices provided by the Bootstrap framework. The positive colour scheme indicates the successful completion of a task and was used for buttons such as the

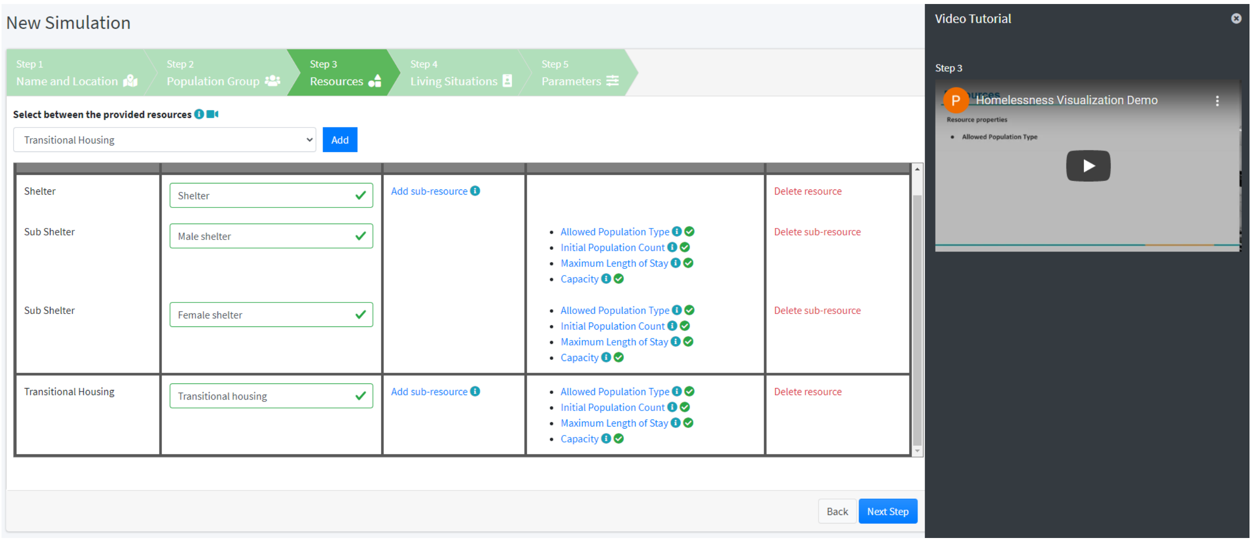

Figure 3.12: The video tutorial slider window. The slider window toggles when the user clicks on the video information icon.

create simulation button and the *complete* indicator badge.

The primary colour is used for buttons and links that help the user to proceed to the next page and adding new elements such as a population type, a resource, or a living situation. The negative colour is used to indicate actions that can have an irreversible effect such as deleting a simulation or receiving an error while making a simulation. The secondary colour is used to show pending status or buttons such as the *Back* button. The information colour is used to indicate that a button or a link will contain relevant information to aid the user's progress.

3.4 Evaluation

We conducted a user study to evaluate and improve the system usability, functionality, visual design, structural design, results presentation, and technical language of the HOMVIZ platform. We were particularly interested in measuring and improving three aspects of our platform: effectiveness, efficiency, and satisfaction.

Participants

We recruited 14 individuals, separated into 3 groups to participate in our user study. All of the participants were graduate students from three different universities: Lakehead University (computer science department), McGill University (epidemiology department), and University of Ottawa (psychology department). Graduate students were selected to help us improve our application to a point where in the future we can invite policymakers to use our application. In addition to graduate students, two senior advisors who are considered to be domain experts in the relevant fields of homelessness and social services assisted in the construction of this application. Our rationale for selecting participants from different disciplines was to gather comprehensive feedback on different aspects of our platform. Each session took 60-80 minutes in total and one person in each group received a $50 Amazon gift card following a random draw.

Apparatus

The focus groups were conducted online using the Zoom application. Participants were asked to share their screens while using our application and completing their tasks. The screens of the participants were recorded using the Zoom screen recorder by two separate systems (one as backup) to be studied later. A heatmap of participant mouse movement on our platform was generated using an online tool called Hotjar[1]. Participants were told that they were free to use the browser of their choice while testing our application. Google Docs and Google Forms were used to share task sheets with users and receive their feedback via a questionnaire.

Procedure

Each usability session began with a brief, 20 minute introduction to our research, the HOMVIZ platform, and a tutorial on how to use our application. To build rapport, we asked participants about their interests and current situation as students during the COVID lockdown. This method proved to be effective in putting participants at ease. Next, the participants were asked to complete a scenario. The task sheets was shared with them using Google Docs which contained an instructional text to make a simulation model and a link to the HOMVIZ platform web application. Next, the participants were asked to fill out the questionnaire that was shared with them earlier. The questionnaire contained a standard SUS questionnaire along with two additional open-ended questions to gain more insight into users' impressions of our application. A sample of SUS questionnaire along with the two additional open-ended questions is presented in Appendix A, Table A.1.

The questionnaire survey was followed by a 30-minute focus group discussion regarding the tester's impressions of the HOMVIZ platform.

Data analysis

The questionnaire survey was evaluated using the SUS formula for producing a standardized score. We used an inductive analysis approach [68] to analyze the screen recordings, transcribed audios, and the mouse movement heatmaps. Mouse movements and audio recordings were reviewed by two graduate assistants to identify (1) points of failure or a particular task that participants found challenging, (2) instances where participants used

[1] https://www.hotjar.com/

the provided help to overcome their problems. Both observations were labeled using an open coding approach and then clustered using an affinity mapping approach to uncover common label groups for the mentioned areas.

3.5 Results

The HOMVIZ platform received a satisfactory SUS standardized score of 75.5 (very good) out of 100 based on the 14 SUS questionnaires that were completed and submitted. Figure 3.13 shows that our participants significantly agreed with the positive statements of SUS questions. They similarly, disagreed more with the negative SUS statements. Our application scored highest for usability, indicating that it was uncomplicated to use, and ease of learning, indicating that participants felt the application was straightforward to learn. In addition to the questionnaire, we collected qualitative data which let us discover more challenges faced by our participants. We grouped those challenges into different emerging themes.

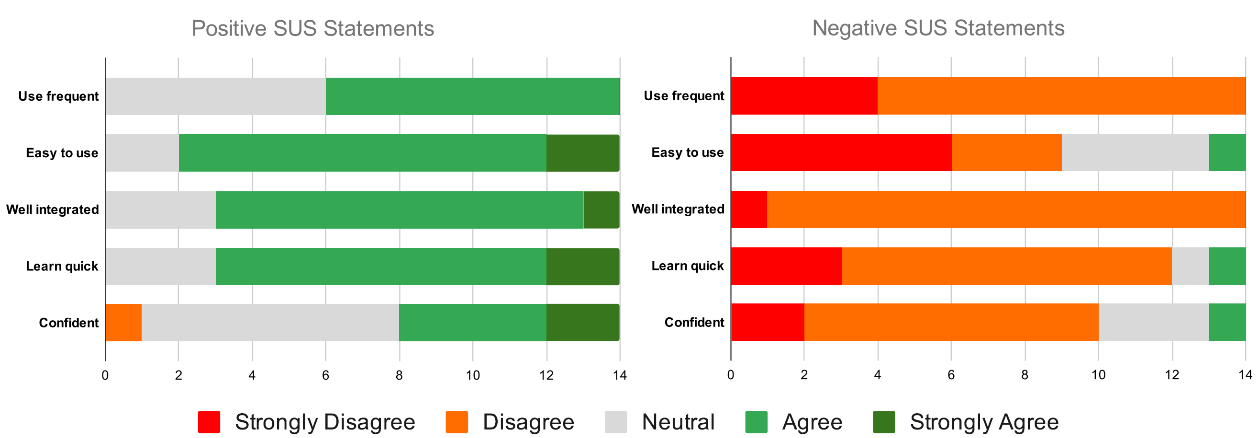

Figure 3.13: Answers to the SUS questionnaire: Positive SUS questions on the left (questions 1,3,5,7,9), negative SUS questions on the right (questions 2,4,6,8,10).

Challenges that were identified

We noticed that the participants with a computer science background were more likely to comment on the UI aspects of the application and participants with epidemiology and psychology backgrounds were more concerned about the definitions and language used on the platform. The following are the themes we discovered by studying the collected qualitative data.

Definitions and incorporated language:

As aforementioned, our simulation algorithm processes a simulation by moving the model's population between resources and living situations in the provided simulation model. We initially named *living situation* as the *state*, consistent with standard terminology used in Markov chain modeling. However, that led to confusion among our participants. The confusion was also observed when looking at the mouse movement of the participants. P04 stated, "*I was not clear on what is the state tab for? Could you help me with that.*" P05 stated, "*I also did not understand what is the meaning of state.*" Given the information, we decided to rename the *state* to *living situation*. Several participants mentioned that our application does not include the gender-inclusive pronoun options in step 2 of the simulation wizard. P01 said, "*Certain categories are not included in the population types. Are there family shelters or shelters for LGBTQ?.*" P02 expressed, "*I know that's how the shelter system works and it's very binary, but maybe approaching it in a way that it includes other population types such as LGBTQ would be better.*" The participants were notified that our algorithm was trained on the data derived from the Montreal site of the *At Home/Chez Soi* project which did not include other gender types.

Functionalities:

We analyzed the raw simulation models created by the participants and compared them to the scenario tasks and found four out of fourteen participants made mistakes adding incorrect population counts. Some of the participants recognized their mistakes before submitting the simulation which was observed in the screen recordings or was asked during the focus groups. P01 stated, "*I did not put the population count in when I selected the population types, and then I went ahead and did the next step but then I got an error when I tried to add the initial population type.*" P06 said, "*I think something is wrong. I cannot add the initial population type. Can you take a look at this please?.*" P09 asked, "*Can I reset the population counts? I forgot to add them on step 2.*" Upon investigation, we found that their error was occurring because they did not realize they had to enter the population counts. To solve this issue, we added a condition to our program where the user must fill in all the population counts before moving on to the next step. This eventually led us to incorporate the same solution in other steps to make sure models are created correctly and to reduce user frustration. To help the user understand their mistake, and guide them in fixing it, we added an error notification dialogue window that shows the relevant error messages along with additional information to fix the problem.

After receiving verbal feedback and observing the user's mouse movements, we learned that users spent a great deal of time and energy reviewing and confirming their earlier inputs. Users complained that the only way to remind themselves of that data was to revisit previous steps in the stepper wizard or to individually investigate each pop-up window input, which was time consuming and frustrating. Figure 3.14 illustrates an example of our observation. The mouse movement heatmap shows that this behavior accounted for 40% of the total clicks on the simulation wizard. P10 stated that "*For me, the big issue was all the data*

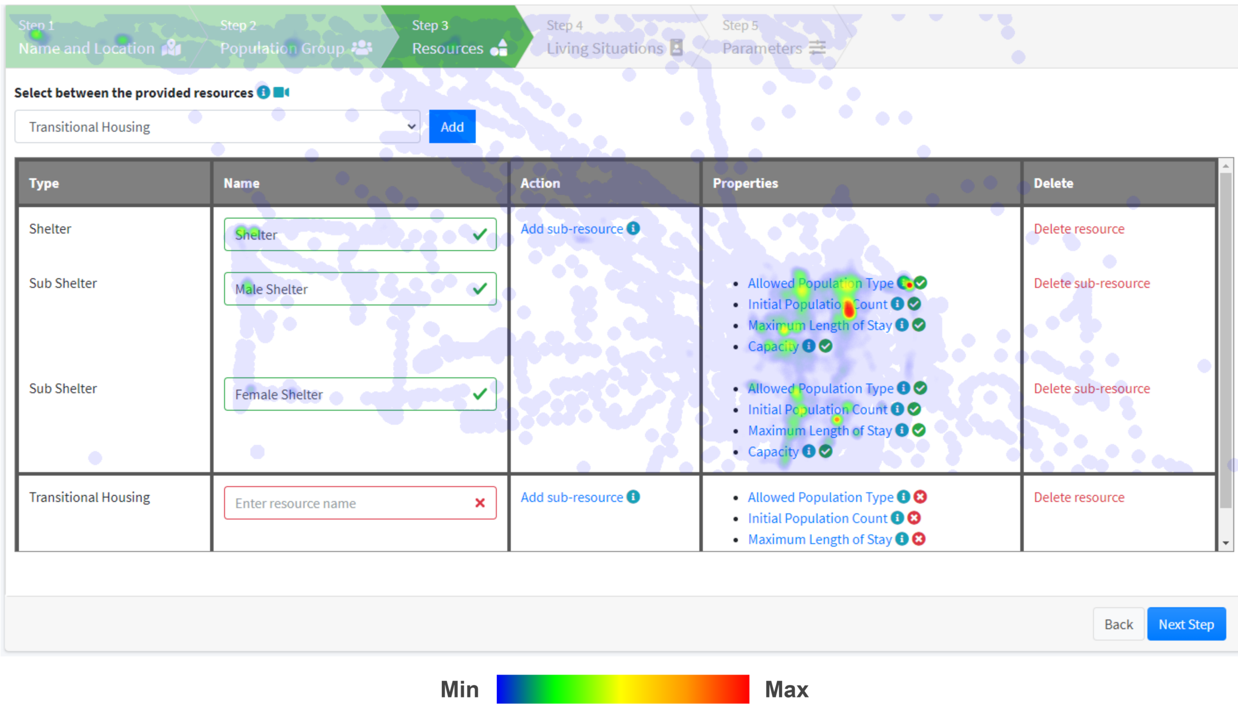

Figure 3.14: A sample of repetitive user mouse movements.

entry boxes were behind those popovers. To access information I have to click on the links. That made it difficult to go back and look at everything I added before." To overcome this issue, we added a new window to the right side of the main window where users create their simulation. The overview window summarizes the user's progress and eliminates the user's need to go back to previous steps and check their inputs. Depending on the user's screen size, the overview window ideally remains next to the simulation wizard and maintains the same height to provide a pleasant aesthetic appeal.

3.6 Observations and Lessons Learned

Based on our evaluation we recommend the following. The web approach functioned as a useful method for this type of software. The web approach enabled our users in different locations and across different platforms and devices to quickly get started and test our application. Our first finding was that the use of a stepper wizard is preferred over a vertical HTML form. Depending on the type of software, if there are too many dynamic and static elements, the form will soon stack up and result in a vertically long page that appears cluttered and can be frustrating to use. Therefore for ease of use, a stepper wizard can be implemented into a platform. The stepper wizard guides the user through the steps of building the simulation, allows them to see the number of steps that are required to finish a task, and does not require excessive scrolling. Additionally, the user-input data can be easily validated and potential errors can be communicated to the user.

When it comes to complex software like simulation and modeling there is a need to provide access to help information. Different methods can be used to present this additional information to the user. It can be presented by having a smaller explanatory text under each element, using an interactive tooltip, adding a tutorial video, and/or using a dialogue window. Showing the information under elements is an acceptable method as long as the platform does not have too many elements on one page; if it does, this method will clutter the page and reduce the efficiency of the platform. Interactive tooltips are an effective way of showing very small pieces of information to the user. Standard tooltips have a smaller font size since the popups appear around the element in the form of a small bubble-like window. As a result of this small pop-up bubble and smaller font, long information shown in the tooltip popover can be hard to read. Additionally, since tooltips mostly appear upon mouse hover they might go unnoticed on touch devices, such as smart phones and tablets, unless a tooltip is programmed to appear upon touch or a click.

The best practice for incorporating tooltips is to attach them to an information icon so that users understand that an action of either clicking or hovering over them is required. Preferably in a hidden slider window, the addition of video tutorials can be beneficial to the user as they provide step by step instructions required for completing the necessary tasks. A Dialogue window or a modal is an effective way of showing longer explanation messages. A dialogue window is a graphical interface tool that appears over the main window and can occupy as much space as is available and can carry additional HTML styling, images, and link anchors.

Based on our experience, dialogue windows are easier to work with and are preferred over other methods for longer information messages and to capture the user's attention. Other styling aspects to look for are the layout width, contrast, text size, alignment, and maintaining a uniform colour scheme. The easiest way to create an appealing layout is by using the freely available Bootstrap framework. It is also important to ensure that the application UI fits the user's device horizontally, this approach is sometimes referred to as the fluid layout.

In this study, we learned that an overview page is very important when using a stepper wizard. Even though a stepper wizard is a useful method to improve the usability of a platform, once the user has moved on to a new step they cannot see the previously added information. Therefore, an overview page that summarizes the overall activities of the stepper is very helpful to users and increases the efficiency and usability of the platform. Strict client-side validation also emerged as necessary. The client-side validation ensures that the user completes the steps of the application correctly and minimizes any potential frustration, confusion, or disappointment.

3.7 Conclusion and Future Work

In this chapter we presented HOMVIZ, a GUI designed to allow policymakers and researchers to easily use a deep learning algorithm to create simulation models. To the best of our knowledge, this research and implementation is the first that reports on the development of a platform that allows policymakers to simulate the effects of implementing social policies. We discussed the important system aspects that we considered while developing the HOMVIZ platform such as robustness, functional utility, and aesthetic appeal. Through brainstorming, usability sessions, and focus groups with three senior professors and 14 participants from multidisciplinary backgrounds, we improved the functionality, system usability, and aesthetic appeal of the HOMVIZ platform. Most participants gave positive feedback regarding the usefulness of the information provided through the platform. The questionnaire survey shows that our application scored highest for usability and ease of learning. The final results show that users can create deep learning simulations using the HOMVIZ platform with minimal help from a technical assistant, thus confirming that it is possible to hide the complexities of an AI algorithm for a social system using a simple to use web-based interface.

In future work, we hope to expand both our sample size and demographic. Based on previous research conducted by Lewis, [30] the SUS questionnaire provides accurate feedback when the sample size is at least 12 or more. While we believe that our usability sessions provided accurate results, we would like to conduct more usability sessions to further improve our application. Moderated usability sessions are time-consuming and hard to be scheduled for a large number of participants. We know that unmoderated usability sessions are easier for a larger number of participants. Next chapter, chapter 4 will compare the moderated and unmoderated usability testing sessions in terms of data reliability and effectiveness. In order to eventually realize that goal, we created a tutorial video for the HOMVIZ platform[1] along with specific task sheets.

The final concern of this study was using graduate students as participants. This study was an initial step towards perfecting the HOMVIZ platform and although our team consisted of domain experts in multidisciplinary fields, we would eventually like to invite policymakers as our participants to experiment with our software in these sessions.

[1] https://www.youtube.com/watch?v=okcfxtIIQdU

Chapter 4

Comparison of Moderated and Unmoderated Remote Usability Sessions for an AI-Supported Simulation and Modeling Software: A Randomized Controlled Trial

All of this chapter is submitted to a peer-reviewed journal as:

- Khoshnevis, P., Tillberg, S., Latimer, E., Aubry, T., & Mago, V. (2020). Comparison of Moderated and Unmoderated Remote Usability Sessions for an AI-Supported Simulation and Modeling Software: A Randomized Controlled Trial.

This work is an extension of the previous chapter and compares the usability testing results acquired in moderated and unmoderated sessions. This chapter investigates the relationship between the aforementioned usability testing sessions and various methods that can be used to measure the quality of participants' responses. A number of inventive methods were used to capture trends of careless responding and other forms of unreliable data.

Keywords: User studies; Usability testing; Usability evaluation methods; Careless responding; Insufficient effort responding;

4.1 Introduction

In the previous chapter, we conducted a moderated survey study to measure the user's perception of the HOMVIZ platform's usability. During our research, we learned that moderated surveys are difficult to schedule, require a significant organizational effort, and can be limiting if a survey requires a large participant sample. The large participant sample, is particularly limiting if the survey requires participants from different demographics and geographical locations. However, related research shows that unmoderated surveys are easier to manage, more efficient, and can include a larger sample size since users are invited to complete a study on their own time while using an asynchronous communication medium such as email. In this chapter, we will compare a moderated and an unmoderated survey for a similar system in order to evaluate data quality for different survey methods. Additionally, this study aims to measure if unmoderated surveys are as effective and reliable as moderated surveys.

Usability surveys and subsequent user interface (UI) improvements are a vital part of the design and development process when establishing effective user interfaces [69]. Usability surveys are conducted in order to measure the users' subjective assessment of a particular UI design [29]. Various aspects of UIs, such as system robustness, functional utility, and aesthetic appeal, can impact the usability experience of an application [61], [47].

Usability surveys are useful for determining what users enjoyed about a particular application as well as what they believe could be improved upon. There are multiple methods and modes in which a usability survey can be conducted and these decisions are based on the available resources, the type of the software, and/or the availability of users [70]. Usability surveys can be conducted after presenting software in moderated or unmoderated modes. Moderated sessions usually take place in a lab or office setting with the help of a moderator to facilitate the study. In contrast, unmoderated sessions can be conducted locally or remotely and do not require the presence of a moderator [71].

The distinction between local and remote studies both for moderated and unmoderated sessions is important to understand. Local moderated or unmoderated studies are conducted in a lab-based setting where all users who are participating in the study are doing so in the same physical space and at the same time [72]. As a result of expanding internet resources, remote studies can be conducted from anywhere in the world and do not require that participants be located in the same region [73]. When compared to local moderated studies, remote moderated studies are generally more cost-effective, less time-consuming, and allow for data collection with a larger sample [74]. Similarly, remote unmoderated sessions are easier to plan, are more flexible, and are more cost-effective than most moderated sessions [75], [76].

Some studies warn that remote usability surveys may miss some contextual information while collecting data from participants [74]. For instance, some researchers have noted a phenomenon named the **mode effect**. The mode effect claims that when similar studies are conducted using different survey modes, the results from these similar studies can vary greatly [75]. Other evidence suggested that remote surveys provide inaccurate results if the surveys are not carefully planned and pretested. For example, online surveys have been found to suffer from a source of error called **careless responding** (CR) where users intentionally or unintentionally respond in a manner that does not reflect their true experience regarding a system [77].

A few attempts have been made to compare the results of moderated versus unmoderated usability surveys for web-based software [78], [79], [71]. A great deal of the existing research that has been conducted regarding moderated and unmoderated survey modes has focused on web platforms whose operation requires little knowledge from participants. Examples of such platforms include news websites, video and music playlist, and e-commerce websites. Absent from those studies are surveys for artificial intelligence (AI) supported web simulation and modeling software which are far more complex to operate. The benefits of AI decision support systems have proven themselves to be valuable and with recent advancements in the computational intelligence field, it is expected that AI-supported software will soon replicate human decision-making skills [80]. Therefore, it is important to study the usability of such complex systems as it is likely that the number of AI-supported software will continue to increase. Web simulation and modeling software differ from other types of web platforms in that they require some prior training where the user can learn about the concepts behind the tools in question. Throughout this prior training, the users can become confident in their understanding of the different variables and parameters required to create a simulation model and run a simulation using the software.

Given the complex nature of such software most usability sessions are conducted with the help of a moderator who initially demonstrates a walkthrough of the application and ensures participants are confident in using the system [34], [33], [35], [81], [32]. This chapter seeks to investigate whether or not unmoderated usability sessions can produce equally useful data and provide users with all of the tools necessary to successfully create a simulation model and run their own simulation. To find answers to this question we have conducted a randomized controlled trial presenting our HOMVIZ platform on two groups: a remote moderated group and a remote unmoderated group. Through our study, we collect various types of data which helps us compare the effectiveness and the level of reliability for each type of session.

It is noteworthy that at the time of this research, the world is currently navigating the COVID-19 pandemic. In Canada, where this research is being conducted, non-essential in-person meetings are strongly discouraged. While the original goal of this research is to investigate the potential of unmoderated usability surveys to increase data output and research-related productivity, it is also being done at a time where the relevancy and necessity for remote modes of communication are at an all-time high.

Research Questions and Contributions

It has been noted that usability studies involving simulation and modeling software are best conducted in moderated environments by a professional who can help users throughout each step. However, research has also shown that unmoderated studies are more cost-effective and allow researchers to include larger sample sizes of participants with varying technological backgrounds and who are from different geographical locations [73]. There are two main questions that have guided our research:

> RQ1: Are the outcomes of the moderated and unmoderated usability testing sessions comparable to one another for a complex simulation and modeling software?
>
> RQ2: What measures should be taken before a remote unmoderated study to ensure data reliability?

Our contribution is three-fold: (1) a detailed explanation of the methods used to conduct the usability testing sessions; (2) a combined analysis of qualitative and quantitative usability data; and (3) freely available code repository of our software and testing methods that were used in this survey[1].

4.2 Related Work

In the following section, we will briefly describe different usability modes, methods, and other elements that must be considered before conducting a usability survey.

[1]https://github.com/pedramvdl31/HOMVIZ

Moderated and unmoderated usability testing

Most moderated usability sessions involve users being guided through their use of a platform or program by an individual who is able to address their immediate concerns and help them work through problems that interfere with their ability to complete their assigned task [78]. Moderated usability sessions, until more recent years, have been one of the preferred methods for testing system usability [79]. However, due to their time, cost, and scope benefits, unmoderated usability evaluation methods have grown in popularity. Unmoderated usability evaluations have the potential to increase the scope and sample size of an experiment by eliminating the need for the presence of developers and researchers in a moderated setting [78]. Additionally, the cost of performing unmoderated usability evaluations is significantly less [71]. A common concern regarding unmoderated usability evaluations is whether or not they will yield accurate and useful information to the same extent as a moderated evaluation. Hertzum et al. [71] tested the reliability of unmoderated usability tests and found that the results observed in both moderated and unmoderated usability evaluations did not differ from each other significantly. Additionally, they investigated whether or not "the evaluator effect" was more prevalent in one method of evaluation over the other. The evaluator effect, while first noted by Jacobsen et al. [82], is described by Lewis as, "the possibility that usability practitioners might be engaging in self-deception regarding the reliability of their problem-discovery methods" [78]. This means that it is believed that depending on the particular evaluator, the results of what usability issues are being discovered will vary across usability testing sessions [83]. Hertzum et al. [71] set out to investigate whether or not moderating a session impacted the presence of the evaluator effect. They reported that the evaluator effect was present to a similar extent in both moderated and unmoderated usability evaluation sessions. Therefore, they concluded that the use of one method over the other would not significantly impact the evaluator effect in a usability study. These findings combined with the significant increase in convenience that unmoderated evaluations provide to researchers and developers indicate that they should be considered as a viable option for testing system usability. Our study will complement the referenced literature by investigating the differences between moderated and unmoderated surveys that require a stage of pre-training.

Usability testing methods

While the focus of this research is investigating the outcome of unmoderated versus moderated usability testing sessions, it is important to consider the methods in which these surveys can be delivered and how their deliveries can impact the results. One of the most frequently researched survey delivery comparisons is that of paper-based versus online delivery. There are a number of reasons why a research team might prefer a web-based approach to deliver surveys [70]. Web-based surveys are more cost-effective, easily distributed, easily tracked and traced, have the capacity to extend their reach across the world, and are generally considered to be more accessible [70]. However, merits to the paper-based survey can be found as well, especially when the “digital divide” of certain test groups is taken into consideration. The digital divide refers to the varying access that individuals have to computers and internet services and how different demographic factors such as age and socioeconomic status affect that access [70]. For example, older participants who do not have the computer literacy necessary for online surveys as well as those who do not have access to a home computer might prefer paper-based surveys as it enables them to participate in research projects.

Ball [84] investigated the feasibility of unmoderated sessions and discussed some of its disadvantages. One disadvantage that is noted is the inability to follow up with participants regarding difficulties that they have encountered with a system. If a moderator is present to follow up with participants to clarify their concerns, comments, or complaints, then it is much easier to ensure the data that is being collected is accurate and reliable. Additionally, the increased use of remote unmoderated sessions may bias the results by failing to include those who lack the technological equipment or knowledge to participate. Additionally, an overrepresentation of certain groups might occur for those whose access to remote unmoderated surveys are hindered [84].

It is noted that the success of an online survey depends greatly on the participants' comfort and experience level with computers [85]. Given this information, we included a computer usage questionnaire in our survey to measure the participants' computer usage frequency and their experience level. Additionally, concerns for one's privacy have been found to affect the usefulness of online surveys with users often distrusting online surveys which require the use of any of their personal information such as name, emails, addresses, etc. [70].

There is conflicting information regarding the honesty and effort that is more likely to result from web-based versus paper-based surveys. Some argue that due to the potential anonymity of web-based surveys, the potential for a more genuine expression of emotion and thought is greater during a web-based survey [86]. However, others argue that the potential for CR in online surveys is greater, and thus it is possible that online surveys may produce more disingenuous data [86].

Careless Responding

Due to the lack of supervision that is experienced during an unmoderated study of any nature, it is expected that some participants might intentionally or unintentionally respond to the survey in a careless or inattentive manner. This phenomenon is referred to as Careless Responding (CR) or sometimes insufficient effort responding [87]. There are several hypothesized explanations for CR as well as different proposed methods for minimizing its effects on survey data. McKay et al. [87] suggest that the likelihood of navigating CR from participants can be linked to personality traits. They predicted that benevolent traits, such as honesty, humility, emotional stability, extraversion, agreeableness, conscientiousness, and openness to experience, may be less likely to provide unreliable responses whereas those with malevolent traits, such as psychopathy and narcissism, could be more likely to submit CR [87]. Ward and Pond [88] discuss how researchers can limit the occurrence of CR by adjusting the questionnaires. One solution they propose is including self-report items that require participants to rate their engagement. Additionally, data-cleansing can be used with caution in order to filter out CR. Ward and Pond argue that in some cases it is better to omit data, thus increasing the potential limitations of the study, rather than to procure low-quality data that does not accurately reflect the study. Other research suggests that in addition to CR, researchers must be aware of the occurrence of attrition in their data [77]. Attrition refers to participants opting out of the survey before it is completed, however they note that more research is needed to assess how attrition impacts survey data [77]. In our study, we implemented various methods, such as timers and manually analyzing data, in order to observe users' behaviour during the survey to recognize CR and occurrences of attrition.

Survey invitation elements

One of the greatest challenges that researchers face when collecting data via surveys is maximizing the response rate. Particularly for remote unmoderated surveys, where participants are being sent the instructions for survey completion, it can be challenging to motivate participants to engage with the study in order to maximize response rates. Petročič et al. [89] suggest that adjusting the initial invitation to participate in the study can greatly impact the response rate. Specifically, they investigate whether using authority, asking for help or referencing a sense of community in email invitations increase response rates [89]. They found that the most effective means of increasing response rates from email initiations for surveys was by asking for help [89]. Given this information, we included a plea for help and attempted to appeal to the users' sense of community in our survey invitation email.

Although having participants engage with the initial invitation is a crucial first step, maintaining and maximizing response rates goes beyond the first contact. Casey and Poropat [90] suggest that one of the main reasons for higher response rates is the system being tested conforming to classical aesthetics. According to Casey and Poropat, the classical aesthetics construct consists of items that represent many general principles of good design and therefore, may be considered to be aligned closely with traditional notions of aesthetics, such as "orderliness, cleanliness and proportion" [90]. Additionally, they suggest that the qualities associated with a classical aesthetic are more likely to motivate participants to complete a study as they increase the likelihood of an enjoyable experience and promote a positive response [90]. Fan and Yan [91] investigate all aspects of a survey, including development, delivery, completion, and return, in order to determine which steps can be taken to maximize response rate. Overall, Fan and Yan suggest emphasis be placed on the aesthetic quality of the platform as well as the use of participant incentives, such as gift cards, in order to maximize response rates [91]. To maximize user satisfaction and response rate to our survey, we incorporated methods such as introducing aesthetically appealing designs to both the system and participants' task sheets. Additionally, we used a draw for Amazon gift cards as an incentive to increase response rate.

4.3 Methodology

In order to properly investigate our research questions, we conducted an exploratory study that compares the data quality of remote moderated versus remote unmoderated usability testing sessions. Additionally, we implemented a number of safeguard methods to identify instances of CR and other types of unreliable data. In the following subsections, we explain our research methodology in detail.

Preparation components

In order to conduct the usability sessions for both usability modes, several documents were prepared. Two separate email invitations were created to be shared with participants. The invitation emails contained all the necessary information regarding our research and links to additional documents such as the information letter, consent form, and task sheet. The information letter provided additional information regarding our research, usability survey, the purpose of the data collection, the type of information being collected, and our information storage policy. Two task sheets were written for each usability mode, only varying where the mode of the survey would impact the user's actions. The task sheets contained step-by-step instructions to help participants complete their simulation setup (preparation documents are presented in Appendix B).

The HOMVIZ Software and Test Tasks

HOMVIZ[1] is a web-based graphical user interface (GUI) to simulate and predict future trends in homelessness. The system is designed to enable users to run different experiments without being concerned about the complexity of the core deep learning algorithm. The target audience for the HOMVIZ platform is policy analysts and researchers. The deep learning algorithm, BEAUT [51] is designed based on the Markov chain decision process and predicts the probability of individuals moving between different states. In this instance, it predicts the various states of homelessness that individuals could be experiencing.

The HOMVIZ application has standard "*login*" and "*registration*" pages. Additionally, the platform contains a "*homepage*" where the user can view a history of their interaction with the system and view the status of simulations. HOMVIZ also has a "*create simulation*" page where the user can create a simulation model using the provided tools and tips. Finally, the "*view results*" page shows the results of user simulations.

In our survey, each participant was given a task sheet via Google Docs that instructed them on the specific steps they needed to take in order to successfully complete the usability session. The task sheet asked the user to register into our system, log in, watch a tutorial video[2] (applicable only to those completing the unmoderated survey), create a simulation, view the results of their simulation model, and answer the questionnaires.

Respondent Selection

72 participants were recruited from a graduate level computer science class. Their scheduled class time was used to conduct our survey. Our study was pre-approved by the Office of Research Services and Ethics Board at the university in which the study was conducted.

[1] https://www.homviz.datalab.science

[2] https://www.youtube.com/watch?v=okcfxtIIQdU

From the original 72 participants, 9 were discarded either due to technical issues or unresponsiveness to our requests and prompts during the survey session. Participants were randomly assigned to either moderated or unmoderated sessions. The overall participants' sample was 71% Male and 28% Female. 84% of participants were between the ages of 21 to 25, 9% were between the ages of 26 to 30, and 6% were between the ages of 31 to 45. Table 4.1 provides more detailed information about the demographic of our survey participants. To incentivize participants, five individuals received $20 Amazon gift cards following a random draw.

Table 4.1: Overview of participants' demographics

Approach	Age [%]	Gender [%]
Overall (N = 72)	21 to 25: 84.1% 26 to 30: 9.5% 31 to 45: 6.4%	Male: 71.5% Female: 28.5%
Moderated Survey (N = 36)	21 to 25: 83.3% 26 to 30: 6.7% 31 to 45: 10.0%	Male: 73.3% Female: 26.7%
Unmoderated Survey (N = 36)	21 to 25: 84.5% 26 to 30: 12.1% 31 to 45: 3.4%	Male: 69.6% Female: 30.4%

Apparatus

Our usability session was conducted using the Zoom Video Communications application. The participants were divided randomly using Zoom's breakout feature. We used smartlook[1] web application and API to generate heatmaps for our study. Questionnaires were developed within our system using HTML 5 forms. Other features such as tracking timers were developed using javascript functions. All the data was stored in MySQL relational database design for our study. The tutorial video was recorded and edited using OBS studio[2] and shotcut[3] freeware and hosted on YouTube in order to simplify user access. The Google Docs online platform was used to send task sheet documents to participants.

Survey Modes

We divided our sample according to two modes: remote moderated and remote unmoderated sessions. Our participants were randomly divided into 4 groups. Groups A and B attended the moderated sessions and groups C and D attended the unmoderated sessions. The next subsections explain both approaches.

Remote moderated usability testing sessions

The remote moderated sessions were conducted with the help of two research assistant moderators. During each session, we began the conversation with a 15-minute introduction to our system and the research behind the HOMVIZ platform. We also conversed with participants regarding the COVID lockdown to build rapport. Following the initial introduction, we conducted a 15-minute tutorial and a walkthrough of our system. At the end of the tutorial, we asked participants if they had any questions or doubts regarding our application. Based on the finding of Micallef et al. [92] we gave our participants a quick summary or tips of what to look for when testing the application and some general knowledge regarding testing as it has been found to improve the performance of untrained participants. Once we answered all the questions we shared a link to a Google Docs which contained the task sheet document outlining all of the necessary information for participants to complete their simulation and the survey. Participants could raise their hand if they had any questions regarding the tasks and we tried our best to assist them. The final step in our usability

[1]https://app.smartlook.com/
[2]https://obsproject.com/
[3]https://shotcut.org/

session was for the participants to answer the questionnaires by visiting a link in the task sheet. Participants were given permission to leave after completing the questionnaire and the moderators stayed present until each participant had finished.

Remote unmoderated usability testing sessions

The Zoom breakout rooms for the unmoderated sessions had two graduate students present as assistants in case of severe technical difficulties. Participants could not ask any questions regarding the system and could leave at will. Our student assistants sent an invitation link to each participant which contained all the information regarding our research and a task sheet. Participants were to follow the task sheet to complete their survey.

Questionnaires

We used the SUS questionnaire along with two additional open-ended questions. The SUS questionnaire is a 10-item questionnaire that uses a 5-point Likert scale ranging from strongly agree to strongly disagree. The SUS questionnaire is the most commonly used usability testing questionnaire [30], [93]. In order to capture additional qualitative feedback from users, two additional open-ended questions were added to the standard SUS questionnaire. The open-ended questions asked participants to express both their negative and positive experiences regarding any aspect of our system. A sample of SUS questionnaire along with the two additional open-ended questions is presented in Appendix A, Table A.1.

Since using our software requires a prior computer and web knowledge, we hypothesized that the participants' potential inability to use our system could result in user frustration affecting their perception related to user satisfaction and thus affect their SUS questionnaire results. Therefore, we used the Computer Usage Questionnaire (CUQ) [94] to capture participants' computer skills for consideration when collecting SUS results in order to ensure their reliability. The CUQ is rated on a 5-point scale (never, rarely, sometimes, often, and very often) and questioned participants regarding their frequency of usage of computer applications (e.g. Microsoft Excel) and the frequencies of activities related to computers (e.g. Skype). A sample of CUQ questionnaire is presented in Appendix A, Table A.2.

Data Analysis

Participants start the usability sessions by answering a short demographic survey on the registration page. The demographic survey was used to determine the demographic characteristics of participants. Standard SUS and CUQ formulas were used to produce standardized scores for both questionnaires. Additionally, the SUS answers were further analyzed to reveal additional characteristics such as:

1. *Use frequent*, derived from questions 1 and 2 which indicate how often the user would like to use our application in the future.
2. *Ease of use*, derived from questions 3 and 4 which measure the user's perception of the usability of the system.
3. *Well-integrated*, derived from questions 5 and 6 which measure the user's perception of the integration of various functions within our application.
4. *Learn quick*, derived from questions 7 and 8 measuring the user's perception of how easy it was to learn how to use our application.
5. *Confidence*, derived from questions 9 and 10 which measure the user's perceived confidence when using our application.

The qualitative data from the open-ended questions were analyzed separately and the answers were labeled using the open coding approach and grouped into 4 categories: *UI related*, *Navigation and functionality*, *Other*, and *Not Answered*.

Multiple mechanisms were programmed in order to detect instances of CR. Most notably, we programmed timers within multiple pages of our application. Timers are program scripts that measure the interval time between the start and end of a task. The timers were placed into *create simulation*, and *questionnaire* pages. The timers were programmed to activate when the page was fully loaded and after the user had clicked on the start button. The data collected demonstrates how long each user was on a particular page and provided insight as to which participants did not pay attention to the instructions and carelessly completed the survey. A separate watch-time timer was added to the tutorial video. The watch-time timer was programmed using the YouTube API in order to record the actual watch time of the tutorial video for each participant of the unmoderated sessions.

The simulation model that was created by participants was stored in the database as a raw JSON file. To further evaluate the participants' understanding of our system and the CR, we manually analyzed the JSON files. Two methods were used to evaluate the raw data. First, we compared the raw data with the task sheet instructions as participants were asked to follow specific instructions when creating their simulation. The instructions were added to ensure that the simulations would be created free of error so that the user could evaluate the UI of the software rather than waste their energy contemplating the values they were entering. To compare the raw data and task sheet instruction, we created a mark sheet that graded every task that the user was supposed to follow. Using this method we scored every participant's simulation model. Second, the JSON raw data were analyzed in order to determine if the participant understood the concept of our software and created a meaningful simulation. In this instance, the participant either received a pass or a fail mark. These methods revealed that some participants followed our instructions entirely and flawlessly while others deviated from the instructions but created a meaningful simulation and clearly understood the concept of our system. Additionally, captured heatmaps were studied to analyze participants' behavior, particular flaws in the system, and showed us the usefulness of some elements.

Using additional relational database queries, we were able to make a correlation between the collected data. To measure statistical significance, we used a two-tailed t-test (equal variance) and Chi-square with Yates correction.

4.4 Results

In the following section, we present the analysis of data from questionnaires, timers, raw simulation data, and heatmaps in order to interpret our study results comparing moderated usability sessions to unmoderated usability sessions. Tables 4.2, and 4.3 shows the summary of our results.

Table 4.2: Comparison of moderated and unmoderated sessions on the continuous outcome measures

Measures	Moderated n, Mean (SD)	Unmoderated n, Mean (SD)	t (DF)	p-value
SUS score	24, 75.8 (15)	20, 72.9 (14)	0.63 (42)	0.52
CUQ score	24, 72.9 (8.1)	20, 72.3 (8.3)	0.22 (42)	0.82
The average time taken to make a simulation	28, 817 sec (431)	30, 615 sec (361)	1.93 (56)	**0.05**
The average time taken to answer a questionnaire	24, 261 sec (117)	20, 303 sec (87)	1.29 (41)	0.20

Table 4.3: Comparison of moderated and unmoderated sessions on the categorical outcome measures

Measures	Moderated (N = 36) n (%)	Unmoderated (N = 36) n (%)	Chi-square	p-value
Did not participate	6 (17%)	3 (9%)	0.50	0.47
Created a simulation	28 (93%)	30 (91%)	0.12	0.72
Used the task sheet's instructions when making a simulation	15 (54%)	8 (28%)	3.96	**0.04**
Created a valid simulation but not the one that was assigned in the task sheet	28 (89%)	23 (79%)	0.44	0.50
Answered the questionnaire	24 (80%)	20 (61%)	0.52	0.46

RQ1: Comparing the outcomes of the moderated and unmoderated sessions

In the following section, we present the methods used to conduct both the moderated and unmoderated studies in our pursuit of answering RQ1.

Questionnaires

Among participants of both the moderated and unmoderated sessions, there were some who decided not to complete the questionnaire. The results show that more participants from the moderated sessions completed the entirety of their assigned tasks including answering the questionnaire. In total, 24 out of 30 participants (80%) from the moderated answered the questionnaire, compared to 20 out of 33 (61%) from the unmoderated sessions. The difference in these proportions is not statistically significant (p-value = 0.46).

The questionnaire answering portion of the usability session started with participants being asked to answer the CUQ, which assesses the participants' pre-existing computer usage habits and overall computer skills. We present the mean result for a five-point rating scale. The results indicate that participants from both survey modes had similar computer usage frequencies. The mean score for participants of the moderated sessions and the unmoderated sessions was 72.9 and 72.3 out of 90, respectively. A t-test shows that the difference between these mean scores was not statistically significant (p-value = 0.82). Given the similarity between these results, it was further determined that any differences between the data collected in the moderated and unmoderated sessions were not due to participants' level of familiarity with computer systems.

The SUS questionnaire results suggest that the participants from the moderated sessions were slightly more satisfied with our UI than those from the unmoderated sessions. The mean score for the SUS questionnaire for the moderated sessions was 75.8 (very good) and was 72.9 (good) for the unmoderated sessions. However, the difference between these means was not statistically significant (p-value = 0.52).

Figure 4.1 indicates that participants from the moderated and unmoderated sessions strongly agreed with the positive statements from the SUS questions. Figure 4.2 suggests that participants disagreed significantly with the negative SUS statements. Upon further analysis of the SUS results, regarding the question suggesting that the system was well integrated, the data indicates that the participants from the moderated sessions mostly agreed with that statement. The same was noted for questions regarding high confidence levels while using the system. However, results from the participants of the unmoderated sessions reported high levels of confidence as well. Upon further analysis, the data suggests that the users' capabilities while using our system was lower than initially thought as we discovered participants from the unmoderated sessions had a larger number of errors in their simulations. This might be due to CR or as the work of McKay et al. [87] suggest, might be due to personal traits such as Agreeableness. It is also noted that the participants from the unmoderated session disagreed more with positive statements regarding their interest in using the system frequently, that the UI was easy to use and well integrated, and that they were able to learn how to use the system quickly.

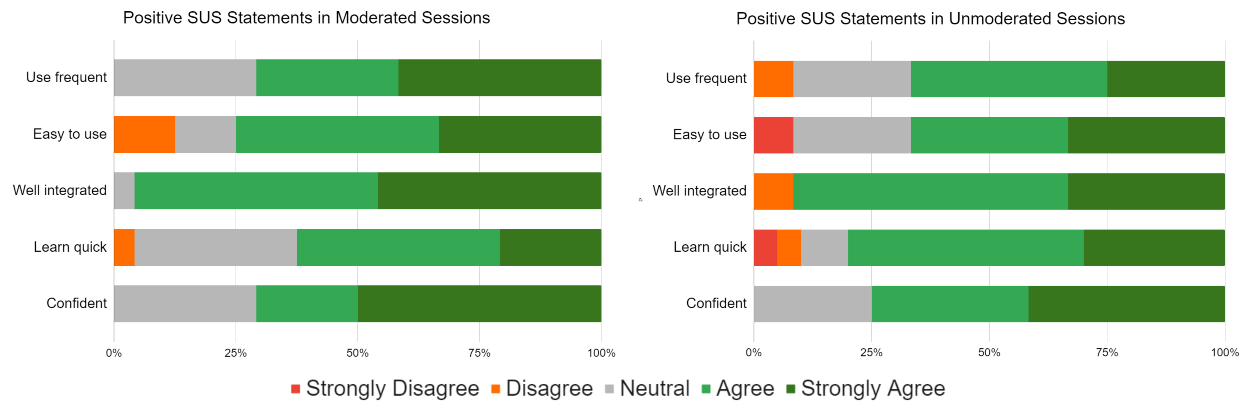

Figure 4.1: Answers to positive SUS questions: Results for moderated sessions on the left and for the unmoderated sessions on the right.

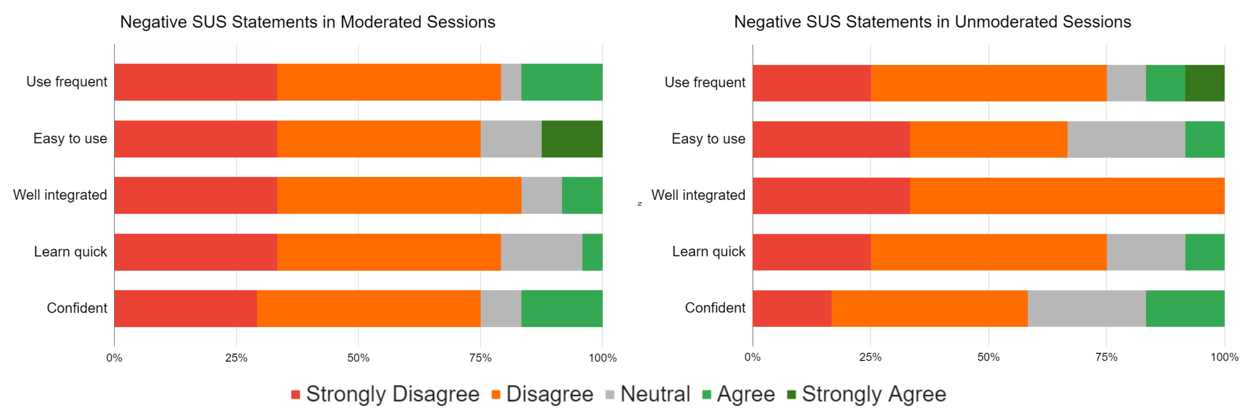

Figure 4.2: Answers to negative SUS questions: Results for moderated sessions on the left and for the unmoderated sessions on the right.

Two open-ended questions were added to the end of the SUS questionnaire. These questions allowed us to obtain qualitative data regarding the platform. Using the collected qualitative data we discovered new recommendations for improving our system and addressing the challenges that participants faced. We grouped participants' feedback into four categories: *UI-related*, *Navigation and Functionality*, *Other*, and *Not Answered*. Figure 4.3 shows our findings. Some participants provided us with valuable constructive comments, for example, *"The pop-up information provided to the user is too complex. Using simpler text can make it more understandable"*, *"It seems like senior citizens will have difficulty using this system. It would be great if there are some accessibility features"*, and *"Since adding properties requires many clicks, it would be nice if after saving one property another property automatically opens"*. The use of open-ended questions was a useful means of collecting qualitative data, however, after analyzing the data we realized the main problem with open-ended questions was the inability to follow up with the participants to clarify their concerns. For example, we received a number of comments that indicated that participants could not clearly explain their feedback due to a lack of domain knowledge. Often their responses were unclear or poorly written, which made understanding the participant's point of view very difficult. Therefore, had the opportunity for following up with participants been available, it would have enabled us to understand the participants' concerns more clearly. 46% of participants from the moderated session and 40% of participants from the unmoderated session did not provide any negative feedback regarding the usability of our application. This may be due to the fact that they genuinely had no negative feedback, that they were not trained to critique the UI design, or that it was due to instances of CR. Had there been an opportunity to follow up with participants either in a focus group or by other means, perhaps more clear and meaningful feedback could have been obtained.

Heatmaps

The heatmap data was collected for all five steps/tabs in the *create simulation* page. Figure 4.4 shows some of our observations in regards to heatmaps generated from the participants' mouse movement. The heatmap data shows that the participants from the unmoderated session clicked on titles and incorrect links more frequently than those from the

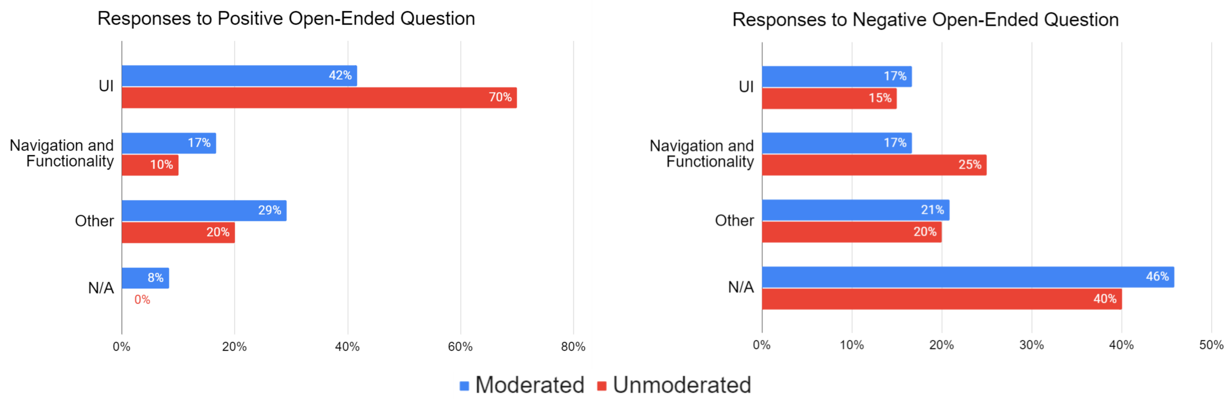

Figure 4.3: Answers to the SUS open-ended questions divided into four categories: UI, Navigation and Functionality, Other and Not Answered.

moderated session. Since we do not have the user's individual mouse movement we cannot conclude if this behaviour is due to CR. It was also observed that the participants from the unmoderated session hover (and potentially clicked) on the various tip links such as popup links, video-tip links, and on the button to close the video-tip slider. This is confirmed with the data we collected using click listeners that counted participants' clicks on the tip links. It can also be inferred that participants from the unmoderated session deleted more rows of data they added, perhaps due to the realization of mistakes or uncertainties. It was also noted that the overview page, the purpose of which was to summarize the user's simulation to minimize unnecessary toggling between steps, was used often.

RQ2: Ensuring data reliability in unmoderated usability sessions

In the following section, we discuss the methods used to ensure data reliability for both the moderated and unmoderated usability sessions.

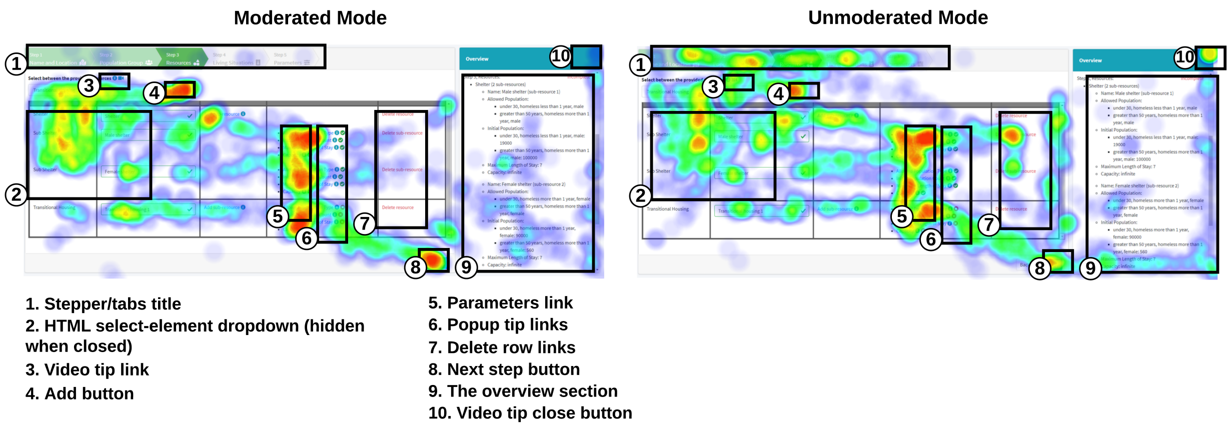

Figure 4.4: Example of the observed mouse movement heatmap. Warmer colors represent a larger mouse movement overlap.

Timers

For further analysis and as a means of assessing the CR, we computed the average time it took participants to complete both the simulation and the questionnaire. The average results indicate that participants from the moderated session created a simulation in 817 seconds and participants from the unmoderated session created a simulation in 615 seconds. Our data shows that a higher percentage of participants in moderated sessions followed the task sheet in comparison to the unmoderated participants (54% and 28% respectively). Therefore, we believe moderated participants might spend more time making a simulation as they were reading and following the instructions. The chi-square result for the simulation completion timer that is presented in Table 4.3 find this difference to be statistically significant (p-value = 0.046).

Participants from the unmoderated sessions took slightly longer to complete their questionnaires than those from the moderated sessions (303 seconds and 261 seconds respectively). The t-test did not find this difference to be statistically significant (p-value = 0.20).

On the provided task sheets, the unmoderated session participants were notified that watching a nine-minute tutorial video was necessary in order to confidently use our system. Our data shows that only 64% of participants watched the tutorial video and only 42% of participants watched the video for more than 300 seconds. The low rate of engagement with the video could be due to its inability to captivate the participants, but participant CR could also be a contributing factor. The average SUS scores from participants who watched the tutorial video for more than 300 seconds was 75.2. This value was very close to that given by participants from the moderated sessions who were given a live tutorial of the system instead of a pre-recorded tutorial video. These findings suggest that having a better understanding of a system could minimize user frustration and result in higher user satisfaction.

Based on the data collected from the open-ended SUS questionnaire answers and additional timer data, we observed that the video tip slider integrated into the platform was effective in helping participants clarify potential misunderstandings. Participants in the unmoderated session used the video slider more often than those from the moderated sessions. The video slider icon was clicked on 21 times by unmoderated session participants and only 5 times by moderate session participants.

Raw data from simulation

While the percentage of simulation turnout between moderated and unmoderated sessions was similar (93% and 91% respectively) our results suggest that participants of the moderated session were more accurate in following the task sheet while completing a simulation. Based on our calculation of the mean score of the rate of accurate simulation completion, 54% of participants from the moderated session did so while 28% of the participants from the unmoderated successfully and accurately completed the simulation. The raw simulation data was further manually analyzed to either pass or fail a participant's attempted simulation. The pass and the fail decision was reached based on our assessment of the users' understanding of the system and the correctness of their simulation. Based on our data, 89% of the participants from the moderated session and 79% percent of the participants from the unmoderated session were able to create an adequate simulation.

Limitations and Challenges

One of the challenges of our study was that the majority of students participating in our survey were doing so in the Indian Standard Time (IST) zone, however, the survey was arranged based on the Eastern Time Zone (EST). Due to the COVID-19 pandemic, many students were unable to travel to Canada to attend their university and were forced to take their classes from their home country where the time zone varied greatly. This time difference caused many participants to join our session past midnight in their time zone and we presume that the late hours may have contributed to instances of CR. Our system was designed using a relational database and as a result, certain information could be correlated with a particular participant. However, the heatmaps were generated using a third-party tool and in consequence, we could not identify the personal mouse movement of any one particular participant. Due to this, the provided heatmap data may be misleading as it could belong to a user whose data was rejected due to CR. Additionally, our sampling method may have introduced bias by inviting participants only from the computer science domain. It is possible that their prior computer knowledge and high-frequency computer usage may have impacted their perception of our system and their usability experience.

4.5 Conclusion

Our research goals were to compare the data quality collected from remote moderated and remote unmoderated usability testing sessions and to investigate what measures should be taken before a remote unmoderated study in order to ensure data reliability.

Research indicates that online unmoderated surveys will only increase in popularity [84], thus finding methods to discard unreliable data will remain important. In our study, we noticed an increase in instances of CR from participants who were completing unmoderated surveys. However, since this was somewhat anticipated, we were able to implement built-in functionalities for our reference when analyzing data that enabled us to identify and separate those who did not thoughtfully complete their survey. Additionally, similar to the findings of Ward and Pond [88], we observed that the virtual presence of a moderator resulted in less CR.

The final SUS result, after careful analysis of all of the collected data, demonstrated that the usability score given by validated participants from the unmoderated study was similar to the results from participants in the moderated study. It was noted that more unmoderated survey participants dropped out of the study, answered fewer questionnaires, and completed fewer tasks than their moderated counterparts. Therefore, the rate and quality of data from the moderated sessions are higher than that of the unmoderated session. However, similar to the observations made by Ball [84], since unmoderated sessions allow for larger sample sizes and safeguards for detecting instances of careless responding are available, the unmoderated method is also very effective for usability research.

To conclude, while the moderated sessions seem to require less effort in data analysis since there were fewer instances of CR, they did require significantly more effort to conduct due to planning and moderating. While the data analysis for unmoderated studies was more labor-intensive, the usability sessions are less costly and can be easily repeated. After these considerations, our data demonstrate that unmoderated usability sessions are worthy of consideration for data collection.

Chapter 5

Smart City Response to Homelessness

All of this chapter was published in the following peer-reviewed journal [95]:

- Khoshnevis, P., Choudhury, S., Latimer, E., & Mago, V. (2020). Smart City Response to Homelessness. *IEEE Access, 8, 11380-11392.*

This chapter presents an example of how computational modeling can be used to mitigate some of the issues faced by individuals experiencing homelessness. This work provides eight heuristic algorithms that were designed to assign homeless individuals to the best possible temporary housing based on their needs, circumstances, and situations while ensuring the fairness of the assignment.

Keywords: Smart Cities; Homelessness; Social Systems; Greedy algorithm; Local search algorithm;

5.1 Introduction

In previous chapters, we discussed methods of effective UI design and presented an easy-to-use GUI that was designed and developed using those methods. We conducted two surveys to evaluate user's perceived usability of the developed application and to compare the data quality of moderated and unmoderated survey methods. This chapter, though it is not directly related to the two previous chapters, yet it contributes to the body of this thesis by proposing a number of computational models that aim to help resolve various issues experienced in the social sector and improve the homeless individuals' access to shelters and safe housings.

Homelessness and poverty are complex problems affecting different types of individuals with different needs. Governments around the world have created various initiatives to address homelessness for different groups of people [96]. For instance, in 2013, the Canadian government budgeted C$119 million for new approaches that could partially solve homelessness [97]. The main one among these is Housing First, which provides homeless individuals, who so desire, their own subsidized, permanent housing, as well as individualized supports. While current policy direction favours expanding the availability of Housing First programs, in practice emergency shelters, transitional housing, and other types of housing (e.g., substance abuse treatment centers), are likely to remain an essential part of available services for years to come.

Most shelters and housing providers have specific admission criteria that include factors such as gender, age group, sobriety, and Indigenous ancestry. Even for someone with a particular profile (e.g., a 35-year-old non-Indigenous man who can show up reliably in a non-intoxicated state, or a woman who has experienced abuse from her partner), there may be several available housing options in a city. The person will seek admission to one place and then to another if the first is full. Shelter personnel (or sometimes case managers) may help a person access transitional housing from an emergency shelter. However, when this happens, the process will invariably be based on limited information.

The development of low-cost web platforms combined with high computational capability provides an opportunity to make this process much more efficient by reducing trial and error and more quickly matching individuals to housing providers that are likely to be better suited to their needs.

Motivating Scenario

In this subsection, we provide a few examples to illustrate how individuals can benefit from such platforms. Consider the following situations (names have been changed to protect the privacy of individuals):

- Diane is 44 years old and has been homeless twice in her life. Once when she was 23 and again at 30. When she was 23, she was asked by her dad to leave her family house. She was struggling and went to stay with friends. Because of her panic attack, she was asked to leave their house. She was scared, confused and wandering in the street. She was eventually helped and got back on her feet. Her struggles could have been less if she was promptly assigned to the most suitable service provider.

- Olivia made an exhausting two-month journey to a different city. She is often dehydrated, has severe stomach pain, and no place to go. She can benefit from a centralized, easy to use application that can recommend the most appropriate housing provider to assist her with her needs.

- Haley was diagnosed with schizoaffective disorder and also experiences gender dysphoria. Haley's parents forced her out of her house. She is scared of going to a homeless shelter because of the negative comments she has heard about homophobic violence or abuse against LGBTQ2S individuals. Haley does not know this, but there are a few service providers nearby her that specialize in assisting LGBTQ2S homeless individuals.

These vignettes illustrate how a centralized mechanism that appropriately matches diverse homeless individuals to housing providers (taking into consideration factors like gender, age group, sobriety, Indigenous ancestry, and the distance to the provider) could improve the living experience of homeless individuals and play a significant role in their social reintegration. Information gathered from such a system (e.g., shortages of specific types of housing) would also be of great value to policy-makers, who currently often rely only on impressionistic and anecdotal information.

Summary of the Research

Currently, the process by which homeless individuals are distributed among available emergency shelters and transitional housing is highly decentralized. In this study, we propose eight novel heuristic algorithms used to create a suitable one-to-one homeless-to-housing matching. Web or mobile applications using these algorithms can be developed to assist homeless individuals by providing them with options for choosing housing that best matches their needs and circumstances. In addition, implementation of such a system can be used to provide policymakers with helpful system-level data. The Conclusions and Future Work section provides a detailed use case of the Smart Housing Framework.

Throughout this study, we use *weight* as a metric to define the relationship between an individual and a particular housing provider, taking into account the Goodness-of-fit of the assignment. The *weight* is inversely proportional to the Goodness-of-fit of the relationship. The main objective of the proposed algorithms is to produce an assignment solution in which every homeless individual is matched with the most appropriate housing provider. Besides that, we are also interested in decreasing the computation time and increasing the $Fairness\ Index$[1]of the entire assignment solution. The $Fairness\ Index$ is the collective equality of the Goodness-of-fit among all homeless individuals within the assignment solution. Therefore, we will compare our algorithms based on execution time, assignment's *weight*, and the $Fairness\ Index$. The accuracy of the final solution is evaluated by comparison to that of the optimal solution.

5.2 Related Work

Minimizing the weight of the homeless-to-housing assignment is similar to minimizing the quantity that is known as makespan [98]. Makespan is a load balancing problem where m machines are given a set of n jobs. A set of jobs, a_i, are assigned to machine m_i. Machine m_i needs to work for a total time of m_t. This is declared as the load on machine m_i. Makespan is the maximum load on any machine in set M, and the scheduling problem by finding the assignment with minimum makespan is an NP-hard problem[2]. Similar to makespan, we are interested in an assignment solution where every homeless individual is offered a housing provider with minimum *weight* based on a set of constraints. Because of this similarity, we

[1]Jain's fairness index is a quantitative measure that is independent of the number of resources, or in this case the number of individuals or the range of weights. The fairness index represents the "equality" of the assignment's weight that is allocated to each individual. If all individuals get assignments with the same weights then the fairness index is 1, and the system is 100% fair. As the disparity increases, fairness decreases and a system that favors only a few individuals has a fairness index near 0.

will summarize relevant makespan research in this section.

The growing interest in Internet of things (IoT) and the ever-increasing number of users provide expanding opportunities for applying optimization methods to resource allocation [99]. Improvements obtained from optimal resource allocation include energy usage reduction of sensors with limited energy, maximizing bandwidth, maximizing and maintaining the quality of service (QoS), minimizing communication collusion in network systems, and many other desirable functionalities and improvements, which results in minimizing cost and maximizing productivity. The resource allocation problem in computer science is similar to the problem statement presented in this chapter. Therefore, we are interested in studies that address resource allocation problems. In the optimization field, makespan is the maximum load on any machine within a set of machines. Makespan is one of the most critical performance indicators, which has been the center of interest for many years [100]. Assignment with minimum makespan is obtained through optimal resource allocation, waste time reduction, and other methods that result in less energy usage, cost reduction, and productivity advancement [100]. Flow shop scheduling and network models are methods used to minimize the maximum load on a machine and maximize productivity [101], [102]. However, an essential factor is the trade-off between minimizing the maximum load and reliability or accuracy. Bi-objective algorithms can eliminate that trade-off using different optimization methods, such as wind driven optimization (WDO) or particle swarm optimization (PSO) [102], [103]. Heuristic algorithms that minimize the maximum load on a machine can be targeted to specific situations to fulfill a requirement, while still maintaining minimum task completion duration, such as preventive maintenance schedules or parallel machines with limited availability [104], [105]. These heuristic algorithms usually applied to NP-hard prob-

[2] We use the term NP-hard referring to solving any NP-problem (non-deterministic polynomial time). NP-hard therefore, means “at least as hard as an NP-complete problem.”

lems that require an instant solution.

Furthermore, an accurate makespan estimation is crucial. Makespan estimations are usually used as an indication of task completion, after which another task begins. Different Machine Learning approaches have been used to increase the makespan estimation accuracy, such as multilayer perceptron type Neural Network [100], which has proven using simulation to be superior to other methods, such as extreme learning machines, and Support vector machine algorithms, in terms of the regression performance indicator. However, a recent publication by the same authors produces better results using convolutional neural networks [106]. In the next section, we formulate the homeless-to-housing problem and define the objective of this research.

5.3 System Model and Problem Formulation

This chapter considers a system with m homeless individuals and n housing providers, where the set of homeless individuals is represented as $H = \{h_1, h_2, ..., h_m\}$ and the set of housing providers is represented as $S = \{s_1, s_2, ..., s_n\}$. From this point forward, homeless individuals are denoted as h_i, $1 \leq i \leq m$ and housing providers are denoted as s_j, $1 \leq j \leq n$. In this work, the assignment of h_i to s_j is denoted as $a_{i,j}$. This can be formally defined as:

$$a_{i,j} \in \{0, 1\}\,;\ \forall\ i, j. \tag{5.1}$$

$$a_{i,j} = \begin{cases} 1, & \text{if } h_i \text{ is assigned to } s_j, \\ 0, & \text{otherwise} \end{cases}$$

Furthermore, a homeless individual can be assigned to only one housing provider. This constraint is defined as:

$$\sum_{j=1}^{n} a_{i,j} = 1;\ \forall\ i \tag{5.2}$$

The capacity is the maximum number of h_i that can be assigned to s_j. Let c_j be the capacity of housing provider s_j. The capacity constraint is define as:

$$\sum_{i=1}^{m} a_{i,j} \leq c_j;\ \forall\ j \tag{5.3}$$

The weight set W contains a relationship weight for every homeless individual and a housing provider. If $a_{i,j}$ denotes the assignment of the individual h_i to the housing provider s_j, then, $w_{i,j}$ represents the Goodness-of-fit of $a_{i,j}$ as an integer between 0 and 100. A lower $w_{i,j}$ value denotes a better matching between an individual and a housing provider. Our algorithms are programmed to utilize the given weights in order to create a better homeless-to-housing matching assignment. The accurate determination of each assignment weight with regards to the impact it has on each individual is a complex social problem and needs further in-depth research. High accuracy of this measurement will affect how well our algorithms impact social settings. However, since our algorithms are designed to utilize the given weights regardless of how they were decided we will randomly generate the weights for all the combinations of homeless individuals and housing providers.

$$w_{i,j} \geq 0;\ \forall\ i, j. \tag{5.4}$$

Furthermore, set W can be presented as:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} & \dots & w_{1,n} \\ w_{2,1} & w_{2,2} & w_{2,3} & \dots & w_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{m,1} & w_{m,2} & w_{m,3} & \dots & w_{m,n} \end{bmatrix}$$

The given problem consists of a set H (homeless individuals) and a set S (housing providers), where the matching process is subject to the capacity constraint. We approach this problem by introducing a combination of Greedy and Local Search algorithms. The algorithms attempt to solve the problem by performing greedy methods so that the final solution contains the lowest weight for each homeless individual without checking every possible pair of associated homeless clients and shelters. We refer to this objective as minimizing the maximum weight. It is important to distinguish between minimizing the sum of weights of the final solution and minimizing the maximum weight of the final solution. To better explain this approach, let us assume the following scenario: Given $H = \{h_1, h_2\}$, $S = \{s_1, s_2\}$, $C = \{1, 1\}$, and $W = \{[6, 1], [10, 6]\}$. The optimal solution for minimizing the sum of all weights is $A = \{a_{1,2}, a_{2,1}\}$ ($h_1 \rightarrow s_2$, $h_2 \rightarrow s_1$). The sum of weights of this assignment set is 11. However, the *maximum assigned weight* in that assignment set was 10 ($h_2 \rightarrow s_1$), which means h_1 was given a much better housing provider at the expense of h_2. This assignment set is considered to be an unfair solution since weights are distributed unevenly amongst individuals, or there is a large inequality between homeless individuals. A homeless-to-housing assignment is considered to be fair when the final solution has high equality of weights among all homeless individuals. Therefore, the optimal solution for minimizing the maximum weight is $A = \{a_{1,1}, a_{2,2}\}$, where the maximum weight is noticeably reduced compared to the previous solution. This objective can be formulated as:

$$min(max\ w_{i,j}.a_{i,j}) \tag{5.5}$$

The homeless-to-housing assignment is a Combinatorial Optimization Problem (COP), in which the ideal outcome is reached by trying all solutions in the search space to find the best possible arrangement. Such problems are also known as NP-hard problems[107]. We formulate the introduced problem as:

$$\text{objective:}\quad min(max\ w_{i,j}.a_{i,j}) \tag{5.6a}$$

$$\text{subject to:}\quad \sum_{j=1}^{n} a_{i,j} = 1;\ \forall\ i \tag{5.6b}$$

$$\sum_{i=1}^{m} a_{i,j} \geq 0;\ \forall\ j. \tag{5.6c}$$

$$\sum_{i=1}^{m} a_{i,j} \leq c_j;\ \forall\ j \tag{5.6d}$$

$$\text{where:}\quad w_{i,j} \geq 0;\ \forall\ i,j. \tag{5.6e}$$

$$a_{i,j} \in \{0,1\};\ \forall\ i,j. \tag{5.6f}$$

The most straightforward method to find the optimal solution is to try the association of each homeless individual and housing provider. This method is referred to as the brute force method, and it is computationally expensive and time-consuming. In some cases, with a realistic population set, the brute force method takes hours to find the optimal solution. To further expand on this issue, let us consider a scenario with five homeless individuals, five housing providers, and a sum of five capacities in all the housing providers. For this problem, there are 120 unique combination sets. However, the number of combinations grows exponentially as the number of individuals and housing providers increases, or when the housing provider capacity decreases. For example, for 15 individuals, 15 housing providers, and 15 capacity among these housing providers, there are 1,307,674,368,000 unique combinations. Hence, this method is not practical given a large number of homeless individuals in major cities like Toronto, or cities with a smaller population like Montreal, with about 3,000 homeless people on a given night [19].

5.4 Proposed Algorithms

In this section, eight novel heuristic algorithms are proposed to solve the problem defined in (5.6a) to (5.6f) in polynomial time complexity. The first four algorithms are greedy algorithms with minor variations. The other four algorithms are local search algorithms that attempt to improve the solution output of the greedy algorithms.

Greedy Method

The greedy algorithms loop through the homeless individuals set H and give priority to an individual with maximum weight values (algorithms are attempting to minimize the final assignment weights). The prioritized individual is sheltered promptly and then removed

Algorithm 1: Greedy - Part 1

Input: $W : [[w_{1,1}, w_{1,2}, \ldots, w_{1,n}], [w_{2,1}, w_{2,2}, \ldots, w_{2,n}], [w_{m,1}, w_{m,2}, \ldots, w_{m,n}]]$;
$H : [h_1, h_2, \ldots, h_m]$;
$S : [s_1, s_2, \ldots, s_n]$;
$C : [c_1, c_2, \ldots, c_n]$;
Output: A;

$A \leftarrow \emptyset$;
$Algo \leftarrow$ select one of the greedy algorithms ; `//` $STDEV\_Algorithm$, $Median\_Algorithm$, $Minmax\_Algorithm$, `or` $Average\_Algorithm$
while $H \neq \emptyset$ *or* $S \neq \emptyset$ **do**
    $D \leftarrow Algo(W)$;
    $personIndex \leftarrow$ get the index position of $max(D)$; `// find the maximum value in` D
    $shelterIndex \leftarrow$ get the index position of $min(W[personIndex])$; `// find the shelter with minimum weight for` $person$
    $a_{personIndex,\ shelterIndex} \leftarrow 1$; `// assigne` h_i `to` s_j
    $H \leftarrow H.remove(personIndex)$;
    $W \leftarrow W.remove(personIndex)$;
    $capacity \leftarrow C[shelterIndex]$;
    $capacity \leftarrow capacity - 1$;
    if $capacity == 0$ **then**
        $S \leftarrow S.remove(shelterIndex)$;
    end
end
return A

from set H. This process continues until either set H is empty or all the housing providers are at full capacity. The key contribution of this research is the following greedy algorithms:

1. STDEV Algorithm
2. Median Algorithm
3. Minmax Algorithm
4. Average Algorithm

Algorithm 1: Greedy - Part 2

Input: $W : [[w_{1,1}, w_{1,2}, \ldots, w_{1,n}], [w_{2,1}, w_{2,2}, \ldots, w_{2,n}], [w_{m,1}, w_{m,2}, \ldots, w_{m,n}]]$;
Output: D

```
Function STDEV_Algorithm(W):
    D ← ∅;
    for each row i in W do
        stdev ← get the standard deviation of row i;
        min ← get the minimum value of row i;
        D[i] ← stdev − min;
    end
    return D
Function Median_Algorithm(W):
    D ← ∅;
    for i in W do
        mdn ← get the median of row i;
        min ← get the minimum value of row i;
        D[i] ← mdn − min;
    end
    return D
Function Minmax_Algorithm(W):
    D ← ∅;
    for i in W do
        min ← get the minimum value of row i;
        max ← get the maximum value of row i;
        D[i] ← max − min;
    end
    return D
Function Average_Algorithm(W):
    D ← ∅;
    for i in W do
        avg ← get the average of row i;
        min ← get the minimum value of row i;
        D[i] ← avg − min;
    end
    return D
```

Generally, the greedy algorithms prioritize a homeless individual that has a maximum disparity among its relationship weights (weight disparities are calculated differently by each algorithm). A maximum disparity suggests a more significant distance between the weights of best and the worst housing provider for that individual. Therefore, it is preferable to match that individual with a housing provider which carries the lowest weight before that housing provider is at full capacity. Consider the following scenario, $H = \{h_1, h_2\}$, $S = \{s_1, s_2\}$, $C = \{1, 1\}$, and $W = \{[2, 10], [3, 4]\}$. In this case, individual h_1 has a higher weight disparity comparing to individual h_2, and the capacity constraint creates a consequential decision, so that, if h_2 is assigned to s_1, then, h_1 must be assigned to s_2, which is the non-optimal solution. $Algorithm\ 1 : Greedy - Part\ 1$ is explained as follows:

1. Line 1 to 3: Algorithm starts by initializing set A. In the next step, we choose one of the greedy algorithms. while-loop is initiated. The while-loop terminates when set H or set S is empty.

2. Line 4: In this step, we call the selected greedy algorithm function. The greedy algorithm compares all individuals and all the housing providers' weights and returns a value for each homeless individual. These values are used to prioritize one homeless individual in each iteration. The homeless individual is given priority by getting matched to a housing provider that carries the lowest weight (lower weight denotes a better assignment).

3. Line 5- 16: We initiate a new variable $personIndex$. The algorithm finds the maximum value in set D and assign the index position of that value to $personIndex$ variable. We then find the index of the housing provider that carries the lowest weight for that individual from $W[personIndex]$ and assign it to $shelterIndex$ variable. Next, the assignment variable $a_{personIndex,shelterIndex}$ is set to 1. $H[personIndex]$ is removed from

set H. We decrement $S[shelterIndex]$'s capacity by 1. if $S[shelterIndex]$'s capacity is 0, then $S[shelterIndex]$ is removed from set S. At the end of the while-loop set A is returned.

$Algorithm\ 1 : Greedy - Part\ 2$ includes four greedy algorithms as functions. A detailed explanation of those functions is as follows:

- Line 1 to 8: For every row i in set W, the $STDEV\_Algorithm$ calculates the standard deviation of the set $W[i]$ and stores the result in *stdev* variable. Next, the algorithm searches for the smallest weight in the same row and stores the weight in the *min* variable. In the last step, the algorithm subtracts the *min* from *stdev* and stores the result to $D[i]$.

- Line 9 to 16: For every row i in set W, the $Median\_Algorithm$ searches for the median value of the set $W[i]$ and assigns the value to the *mdn* variable. Next, the algorithm searches for the smallest weight in the same row and stores the weight in the *min* variable. In the last step, the algorithm subtracts the *min* from *mdn* and stores the result to $D[i]$.

- Line 17 to 24: For every row i in set W, the $Minmax\_Algorithm$ searches for the minimum and the maximum weights in row $W[i]$ and assigns the values to *min* and *max* variables respectively. In the last step, the algorithm subtracts the *min* from the *max* variable and stores the result to $D[i]$.

- Line 25 to 32: For every row i in set W, the $Average\_Algorithm$ calculates the average value of the set $W[i]$ and stores the result in the *avg* variable. Next, the algorithm searches for the smallest weight in the same row and stores the weight in the *min* variable. In the last step, the algorithm subtracts the *min* from *avg* and stores the result to $D[i]$.

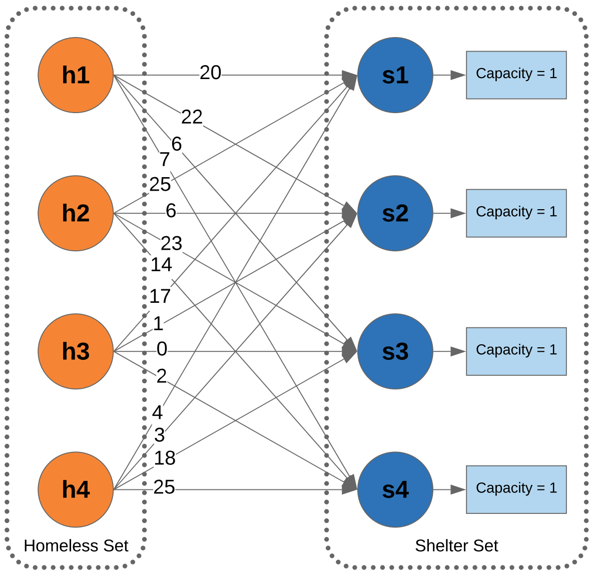

Figure 5.1: Scenario 1

Working Example

In this subsection, each algorithm is executed as a separate program instance, and the step-by-step working example of the algorithms is illustrated and compared using the input data defined in Scenario 1 (Figure 5.1). Scenario 1 is defined as follows:

Scenario 1: $H = [h_1, h_2, h_3, h_4]$, $S = [s_1, s_2, s_3, s_4]$, $C = [1, 1, 1, 1]$, and $W = [[20, 22, 6, 7], [25, 6, 23, 14], [17, 1, 0, 2], [4, 3, 18, 25]]$.

Here we explain the operation of the algorithms:

- Line 1 to 3: Each algorithm starts by initializing set A and begins the while-loop.

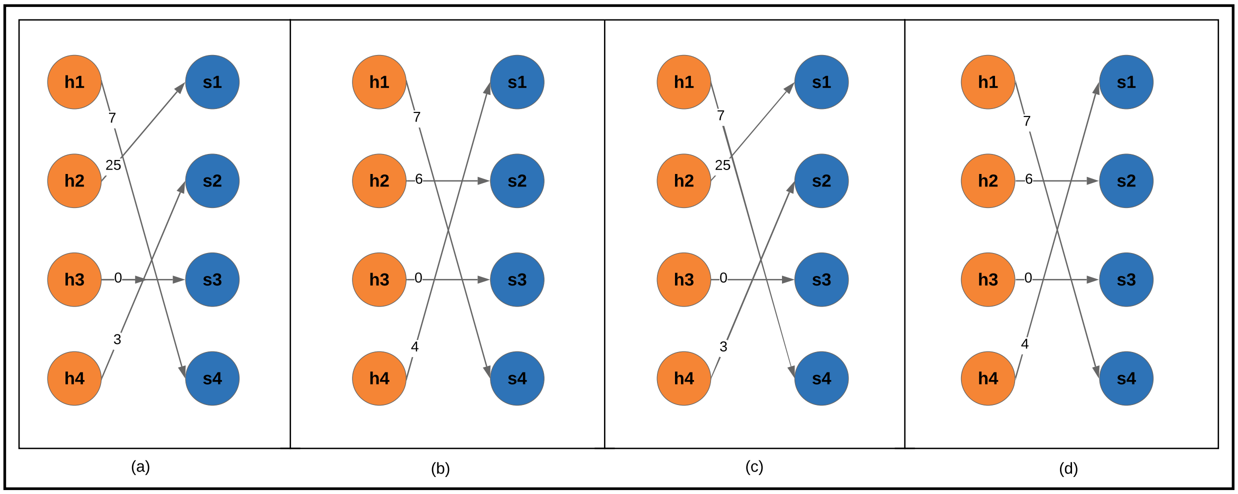

Figure 5.2: Scenario 1 solution. (a) a solution by STDEV_Algorithm [7, 25, 0, 3]. (b) a solution by Median_Algorithm [7, 6, 0, 4]. (c) a solution by Minmax_Algorithm [7, 25, 0, 3]. (d) a solution by Average_Algorithm [7, 6, 0, 4].

- At the next step, any of the algorithms that are defined in $Algorithm\ 1 : Greedy - Part\ 2$ initiates a for-loop for every row i in set W. Within the loop, the dispersion value for each row is calculated. The values are stored in set D. At the end of the loop, the output variable is returned to $Algorithm\ 1 : Greedy - Part\ 1$. The output array is stored in set D. In the first iteration, D contains the following values:

  1. STDEV_Algorithm: $D = [2.4, 2.7, 8.0, 7.7]$
  2. Median_Algorithm: $D = [7.5, 12.5, 1.5, 8.0]$
  3. Minmax_Algorithm: $D = [16, 19, 17, 22]$
  4. Average_Algorithm: $D = [7.75, 11.0, 5.0, 9.5]$

- Line 5 to 16: The index positions in set D are similar to those of set H. For example, the value at index $D[0]$ belongs to the element at index $H[0]$. An individual with the highest corresponding value in set D is given priority and matched with a housing provider promptly. For instance, in the case of $STDEV\_Algorithm$, homeless individual h_3 has the highest value in set D. Therefore, h_3 is assigned to the housing provider s_3, which carries the minimum *weight* for that individual. Next, the assignment variable $a_{3,3}$ is set to true. The homeless individual h_3 is removed from the set H. Row w_3 is removed from the *weight* set W. The housing provider s_3's capacity is decremented by 1, and since the capacity of s_3 is 0, s_3 is removed from the housing provider set S. In the next iterations, the same operations are repeated until either the homeless individuals set H, or the housing providers set S is empty. To avoid redundancy, we refrain from explaining the similar steps in the next iterations; however, the returned values in the next two iterations are as follows:

1. a) STDEV_Algorithm : $D = [1.1, 3.5, null, 9.4]$
   b) Median_Algorithm : $D = [1, null, 2, 14]$
   c) Minmax_Algorithm : $D = [14, 11, 17, null]$
   d) Average_Algorithm : $D = [5.0, null, 6.3, 11.6]$

2. a) STDEV_Algorithm : $D = [2.1, -6.2, null, null]$
   b) Median_Algorithm : $D = [0.5, null, 1.0, null]$
   c) Minmax_Algorithm : $D = [13, 11, null, null]$
   d) Average_Algorithm : $D = [0.5, null, 1.0, null]$

- The last two homeless individuals are matched with a housing provider one after another. Figure 5.2 shows the assignment set A that is returned by each algorithm. The assignment set A is as follows:

  1. STDEV_Algorithm : $A = \{a_{3,3}, a_{4,2}, a_{1,4}, a_{2,1}\}$
  2. Median_Algorithm : $A = \{a_{2,2}, a_{4,1}, a_{3,3}, a_{1,4}\}$
  3. Minmax_Algorithm : $A = \{a_{4,2}, a_{3,3}, a_{1,4}, a_{2,1}\}$
  4. Average_Algorithm : $A = \{a_{2,2}, a_{4,1}, a_{3,3}, a_{1,4}\}$

Worst Case

The main feature of the introduced greedy algorithms is the trade-off between the running time and the solution accuracy. Because of that trade-off, greedy algorithms can produce unsatisfactory results. In this section, we give a few examples that illustrate some of the shortcomings of our greedy algorithms. The $Minmax\_Algorithm$ and the $Median\_Algorithm$ fail to take every weight into account while making a decision. This can be better explained using the following example: Let us assume that the algorithms are given a single row of weights $w_i = [1, 8, 10, 11, 200]$. The $Minmax\_Algorithm$ subtracts the minimum weight ($w_{i,1} = 1$) from the maximum weight ($w_{i,5} = 200$) and stores the value in set D. The $Median\_Algorithm$ subtracts the minimum value ($w_{i,1} = 1$) from the median value ($w_{i,3} = 10$) and similarly stores the results in set D. Both of these algorithms fail to take other shelter weights into account and only measure two values. Similarly, the $STDEV\_Algorithm$ and the $Average\_Algorithm$ fail to give a certain weight higher importance. For example, given a single row of weights $w = [1, 2, 3, 100, 200]$. In this case, both algorithms return a high dispersion value regardless of the second and third suitable housing providers (with low weights). These deficiencies can be efficiently improved using local search algorithms. In the next subsection, we explore a Swap-Based Local Search algorithm to improve the output solution of the greedy algorithms.

Algorithm 2: Local Search

```
Input: W : [[w_{1,1}, w_{1,2}, ..., w_{1,n}], [w_{2,1}, w_{2,2}, ..., w_{2,n}], [w_{m,1}, w_{m,2}, ..., w_{m,n}]];
H : [h_1, h_2, ..., h_m];
S : [s_1, s_2, ..., s_n];
A ;                    // assignment solution (output) of a greedy algorithm
Output: A
flag ← True;
while flag do
    flag ← False;
    MaximumAssignedWeight ← 0;
    for each a_{i,j} ∈ A do
        // this loop returns the maximum weight in set A which we are
           trying to minimize
        w ← W[i][j];                    // get the weight of this assignment
        if MaximumAssignedWeight < w then
            MaximumAssignedWeight ← w;
            VPIndex ← i;                               // person i's index
            VPShelterIndex ← j;                // person i's shelter index
        end
    end
    for each p ∈ H do
        if p ≠ VPIndex then
            pShelterIndex ← j;      // shelter j that was assigned to p
            swapA ← W[p][VPShelterIndex];
            swapB ← W[VPIndex][pShelterIndex];
            if both swapA and swapB were smaller than
             MaximumAssignedWeight then
                A ← swap the vulnerable person' shelter with p's shelter;
                flag ← True;
            end
        end
    end
end
```

Local Search Method

The local search algorithm takes a feasible solution to the problem, which is returned by the greedy algorithms and repeatedly implement small changes to improve the results. In every iteration, the local search algorithm finds a homeless individual who was given the worst housing provider and swap the individual's housing provider with other homeless people in set H. If, after the swap, the *maximum assigned weight* of the entire solution is minimized, the algorithm updates the solution set and repeats the iteration. The local search algorithm has polynomial running time and achieves a substantial improvement, as shown in the Complexity Analysis subsection. $Algorithm\ 2 : localsearch$ is explained as follows:

1. Line 1 to 4: At the first step, we initiate the variable $flag$ with initial value $True$. We begin a while-loop that terminates when $flag$ is $False$. Within the loop, we set the $flag$ to $False$ and initiate the variable $MaximumAssignedWeight$ with initial value 0.

2. Line 5 to 12: We begin a for-loop that runs for each $a_{i,j}$ in set A. Within the loop, we access set A and find the assignment with the maximum weight and assign that weight to $MaximumAssignedWeight$ variable. Similarly, we store the person's index (i) and the shelter's index (j) of that assignment to $VPIndex$ and $VPShelterIndex$ respectively. (this loop returns the maximum weight in the assignment set A that we are attempting to minimize)

3. Line 13 to 15: In the next step, we initiate a for-loop for every element p in set H. Within the loop, if p is not equal to $VPIndex$ (person p is not the same person as the vulnerable person (VP)), then we access A and retrieve the shelter's index (j) that was assigned to p.

4. Line 16 to 17: Next, we access the weight set W and retrieve the weight of the shelter that was assigned to VP for p. Similarly, from W we retrieve the weight of the shelter that was assigned to p for VP. These two weights are stored in $swapA$ and $swapB$ respectively.

5. Line 18 - 20: If $swapA$ and $swapB$ weights are smaller than $MaximumAssignedWeight$ that means the highest weight in the assignment solution A was reduced. If so, we will update the assignment solution with the new shelters and set the $flag$ to $True$.

Complexity Analysis

Algorithm 1: Greedy Part 1 and Part 2: In this subsection, we provide the complexity analysis of the greedy algorithm and the local search algorithm. The worst case running time of $Algorithm\ 1 : Greedy - Part\ 1$ is $O(m^2 + (m * n))$. The breakdown of this analysis is as follows:

- Line 3: $Greedy - Part\ 1$ runs for every homeless individual in set H. Therefore, in the worst case, the running time of $line\ 3,\ Part\ 1$ is $O(m)$.

- Line 4: This line is calling the functions defined in $Greedy - Part\ 2$. We find that every greedy algorithm defined in this research has the worst case running time of $O(m * n)$.

- Line 5: The algorithm searches through set D. Set D contains one value for every element in set H. Therefore, in the worst case, the running time of this procedure is $O(m)$.

- Line 6, 8, 9, 12, and 13: The worst case running time complexity of accessing an array is $O(1)$. Therefore, the running time complexities of these lines are $O(1)$.

- Line 7: The program accesses a predefined row index and searches within that row. The row contains one value for every housing provider in set S. Therefore, the worst case running time of this line is $O(n)$.

- Line 10,11: In the worst case, the running time for a deletion procedure is $O(m)$. Therefore, each line has a running time of $O(m)$.

- Line 15: This line does not run in every iteration. However, in the worst case, this line will be executed in every iteration with the running time of $O(n)$.

Algorithm 2: Local Search: The worst case running time of the local search algorithm is $O(K * m^2)$:

- Line 2: The while-loop requires a complexity of $O(K)$. Where K is the number of iteration of the while-loop. K is defined as $n*(max(w_{i,j})-min(w_{i,j}))$. In the worst case scenario, the greedy algorithm returns a solution set where the *maximum assigned weight* is equal to the $max(w_{i,j})$ (maximum weight in set W). Since the local search algorithm improves the solution set by at least 1 weight in every iteration (for every person), then in the worst case the while-loop runs for $n * (max(w_{i,j}) - min(w_{i,j}))$ times.

- Line 5 $\sim$ 13: Line 5 is repeated for every element in set A. Since the assignment solution possibly contains an assignment for every homeless individual then $|A| \leq |H|$. Therefore, Both lines are executed for every element i in set H. The worst case running time for each of these lines is $O(m)$.

- Line 19: The worst case running time for the array insertion is $O(m)$. Therefore worst case running time of this line is $O(2 * m)$ which requires two insertion.

- Other lines of the algorithm have the worst case running time of $O(1)$.

5.5 Performance Evaluation

Experimental Setup

We perform the experiments on a Windows machine, with an Intel Core i5-8500 CPU and 8.00GB of memory. Algorithms were developed in Python version 3.6.5, and the ILP solver was developed using IBM ILOG CPLEX Optimization Studio [108]. This research is interested in the performance accuracy, run-time, and fairness of the algorithms for large enough population. The algorithms are compared for different population sizes (100, 300, 500, 700, 1000) and randomly generated weights for the problem defined in (5.6a) to (5.6f).

Result Comparisons

The optimal solution for each scenario was obtained using the ILP solver. To examine the accuracy of our algorithms, we compare each of the solutions to that of the ILP solver. Furthermore, the algorithms are compared in terms of the assignment weights. An algorithm with the lowest *maximum assigned weight* is considered a better algorithm. Table 5.1 shows the *maximum assigned weight* (an assignment with the maximum weight) within the solution set which was produced by each algorithm. In terms of the *maximum assigned weight* metric the $Average + local\ search$ algorithm had the best results. Table 5.1 also shows the performance difference between the algorithms with and without the Local-Search procedure. It can be observed that the Local-Search algorithm significantly improved the performance of the greedy algorithms. As discussed in the Complexity Analysis subsection, in the worst case scenario, the Local-Search algorithm requires at least $K = n * (max(w_{i,j}) - min(w_{i,j}))$ iterations to improved the solution set. However, Table 5.3 shows the iteration count of each Local-Search algorithm (until termination) where the number of steps was significantly lower than the worst case scenario. Figure 5.6 illustrates the average accuracy of each algorithm.

Table 5.1: Performance comparison

Algorithms	**n=100**	**n=300**	**n=500**	**n=700**	**n=1000**	**n=1300**
Optimal Solution	25	22	21	21	20	20
STDEV Algorithm	79	83	68	55	81	57
Median Algorithm	40	52	50	42	38	32
Minmax Algorithm	70	76	43	54	76	58
Average Algorithm	51	35	41	40	38	38
STDEV + local search	34	28	25	24	24	26
Median + local search	34	27	27	26	24	25
Minmax + local search	38	27	26	24	24	26
Average + local search	**33**	**27**	**24**	**24**	**23**	**23**

Table 5.2: CPU computation time comparison (HH:MM:SS:FF)

Algorithms	**n=100**	**n=300**	**n=500**	**n=700**	**n=1000**	**n=1300**
Optimal Solution	0:0:02:62	0:15:22:59	02:26:49:37	07:46:33:22	23:41:52:23	38:03:40:78
STDEV Algorithm	0:0:0:13	0:0:3:44	0:0:15:88	0:0:42:60	0:2:6:51	0:4:30:16
Median Algorithm	0:0:0:036	0:0:1:06	0:0:5:16	0:0:15:04	0:0:46:29	0:1:43:18
Minmax Algorithm	0:0:0:013	0:0:0:27	0:0:1:18	0:0:3:18	0:0:9:28	**0:0:21:06**
Average Algorithm	0:0:0:04	0:0:1:12	0:0:5:82	0:0:14:26	0:0:40:55	0:1:26:79
STDEV + local search	0:0:0:1305	0:0:3:48	0:0:15:99	0:0:42:90	0:2:6:98	0:4:30:87
Median + local search	0:0:0:04	0:0:1:08	0:0:5:18	0:0:15:16	0:0:46:80	0:1:43:98
Minmax + local search	0:0:0:0139	0:0:0:31	0:0:1:26	0:0:3:52	0:0:9:92	**0:0:22:07**
Average + local search	**0:0:0:0404**	**0:0:1:13**	**0:0:5:90**	**0:0:14:44**	**0:0:41:02**	**0:1:27:32**

The *Average + local search* algorithm obtained 92% accuracy followed by *STDEV + local search* (91.6%) and *Median + local search* algorithm (91.2%). Using the run-time comparisons that are presented in Table 5.2, it can be observed that our algorithms are significantly faster than the solver's program. For example, in the case of the *Minmax +*

Table 5.3: Iteration count (local search algorithm)

Algorithms	**n=100**	**n=300**	**n=500**	**n=700**	**n=1000**	**n=1300**
STDEV + local search	27	36	41	45	59	77
Median + local search	9	16	12	18	30	35
Minmax + local search	7	44	40	54	55	51
Average + local search	5	10	21	24	34	**42**

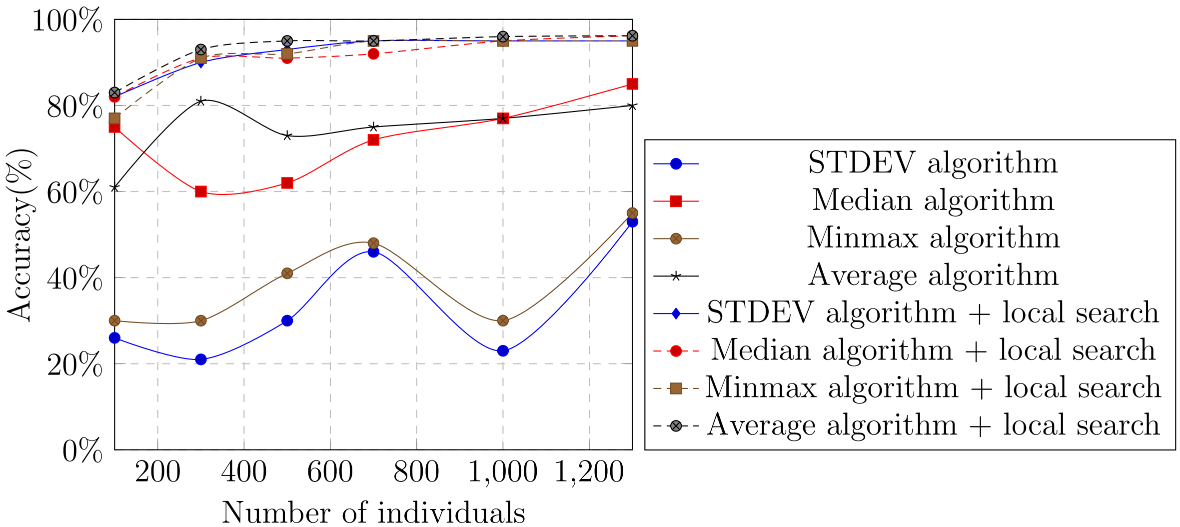

Figure 5.3: Algorithms' accuracy vs the number of individuals

local search algorithm (1300 homeless individuals), the execution time was improved from 136800 to 21 seconds.

To examine the relationship between the number of shelters and algorithm's performance we created several scenarios varying the number of shelters (n = [10, 50, 100, 150, 200, 250, 300, 350, 400]), with a fixed number of homeless individuals (m = 400). For every number of shelters n, we created three scenarios (randomly generated weights and capacities). The capacities were randomly distributed across the shelters. To create a feasi-

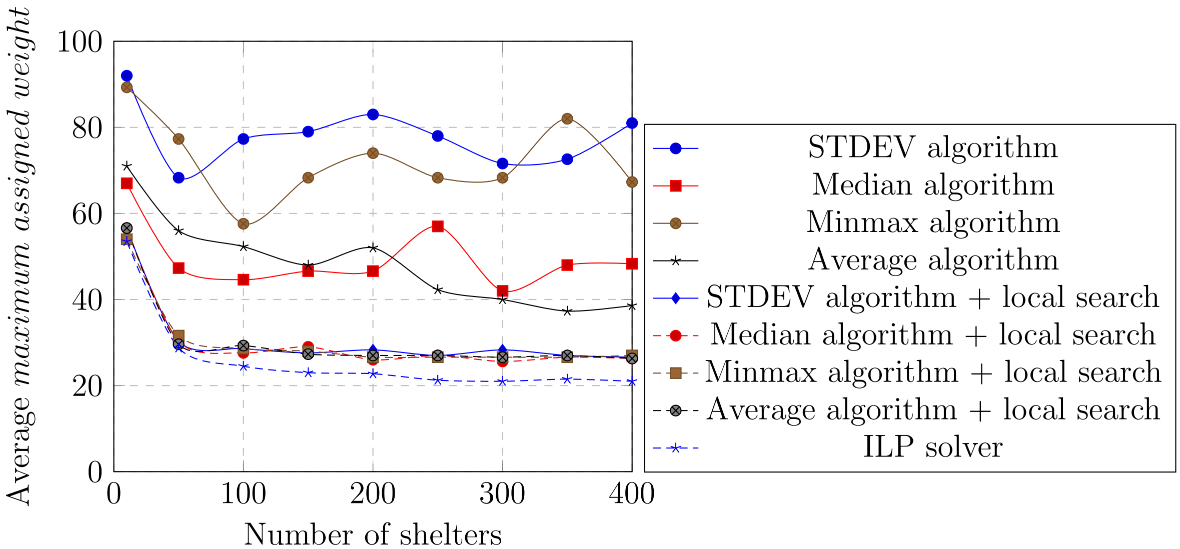

Figure 5.4: Algorithms' performance vs the number of shelters

ble solution the shelter's capacity matched the number of individuals in every scenario. In total, there were 27 scenarios. To compare the results of the algorithms we collected the *maximum assigned weight* of each solution.

Figure 5.4 shows the results of this comparison. It is noticeable that after a certain number of shelters, the performance of the algorithms did not continue to improve. These observations can help in terms of shelter capacity utilization.

Furthermore, fairness comparison was performed based on Jain's Fairness Index [109]. Jain's Fairness Index provides strong feedback on equality of all assigned weights among homeless individuals, but not their magnitude. It is important to note that the optimal solution produced by the solver is not necessarily a fair assignment since fairness maximization was not the objective of this research. Figure 5.5 shows the Fairness Index of the ILP solver and other algorithms. The ILP solver obtained a Fairness Index of 99.90%, followed by the *Average + local search* algorithm (99.75%) and the *Median + local search* algorithm

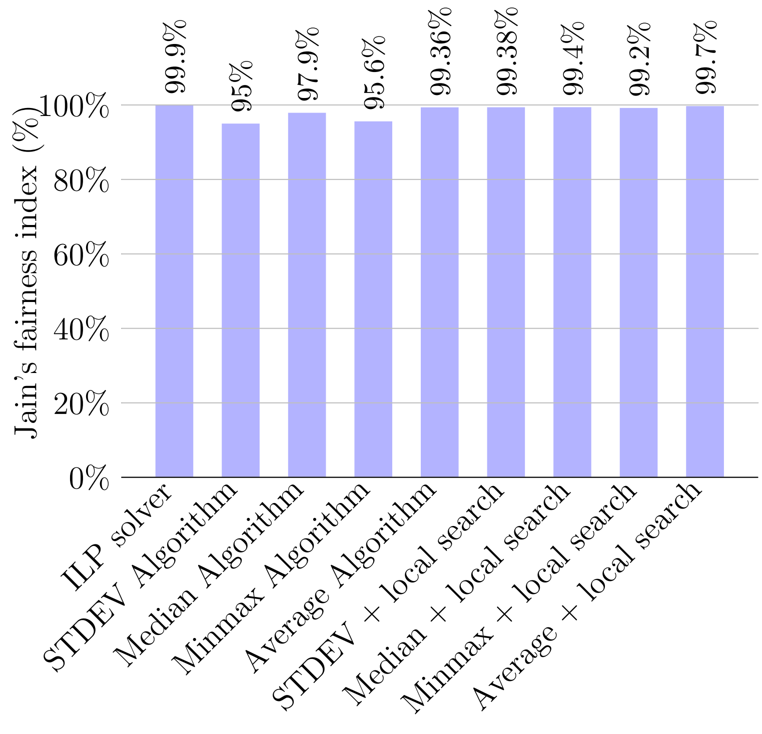

Figure 5.5: Fairness index comparison

(99.49%). Based on all observations, it is evident that the *Average + local search* algorithm performed better than other algorithms. Figure 5.3 presents accuracy comparisons of the proposed algorithms for a different number of homeless individuals. Similar to the previous comparisons the accuracy is measured in relevance to the optimal solution provided by the ILP solver. We can conclude that a larger homeless population did not have a negative effect on the performance of our algorithms. The source code of the experiments can be found in our GitHub repository[1].

[1]https://github.com/pedramvdl31/Smart-City-Response-to-Homelessness.git

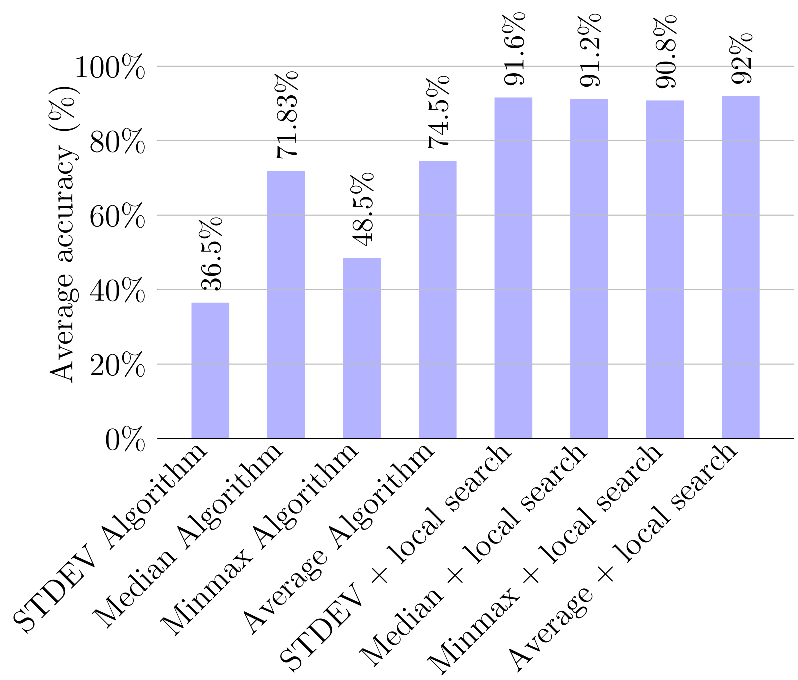

Figure 5.6: Average accuracy comparison

5.6 Conclusions and Future Work

The homeless population is highly diverse. Currently, in Canada, housing providers typically provide different types of services to individuals with different needs. However, the diversity of the population, their geographical dispersion across cities, and other circumstances make the homeless-to-housing matching a complex task. Our works aims to lay a foundation for the development of a platform to facilitate accomplishing it more efficiently.

We believe our proposal is realistic. In the United States, the UK, and at least many countries in continental Europe in addition to Canada, emergency shelters and transitional housing providers are usually owned by a variety of non-profits, which can be large or small. For our application to be realizable, two conditions need to be met: (1) each provider needs to update information about remaining beds in an electronic system in real-time; and (2) information about remaining spaces contained in each electronic system needs to be gathered and made available, also in real-time. The first of these conditions seems to us likely to already be met by many providers. It is in the provider's interest that individuals already registered for the night, or for whom a place is being reserved, be recorded in an electronic system so that: (a) the individual at the front desk can keep track of how many spaces remain available, and of what types, at any given moment; and (b) when the time comes to prepare reports for the government or funders, including information about occupancy rates, etc., these can be quickly put together. The second condition is not, to our knowledge, currently met. However, considering existing integrated online platforms, e.g., in commercial applications, it is quite feasible technically. It is only a matter of government and providers deciding to implement a proposal such as ours. We also note that homeless individuals who do not have a smartphone (surprisingly perhaps, a number do), someone at the front desk of a provider could access the system and be able to quickly and reliably refer the individual to an appropriate resource with available space. We note also that the recent growing interest in smart cities [14] invites the development of such a platform. In this research, we introduced several algorithms that produce suitable homeless-to-housing matching. Besides maximizing the Goodness-of-fit, fairness, algorithms accuracy (in comparison with the optimal solution), and the computational time figured among our objectives. Algorithms that were introduced in this research delivered satisfactory results. Our best algorithm ($Average + localsearch$) produces 92% accuracy, 99.75% fairness, and approximately reduces the computation time by 38 hours. An interesting extension of this work would be a bi-objective algorithm that maximizes the Goodness-of-fit while maximizing the Fairness Index. We are also interested to see the Smart Housing Framework in practice. Smart Housing Framework is potentially a complete platform consists of the assignment algorithm program, a policymaker control application, and a specialized application for homeless individuals. Figures 5.7 and 5.8 are a representation of what we envision for our future real-time application. Figure 5.7 shows the application for policymakers that provides all the necessary tools and statistical data in order to assist the policymakers in the decision making process. Figure 5.8 shows the phone application for homeless individuals who receive notifications upon a new assignment. Additionally, the phone application can assist the individual with the walking direction to the housing provider.

Other extensions of this work and extra features that can be beneficial to policymakers and homeless individuals are as follows:

- The ability to access the third-party API in order to retrieve additional information and a history report about the homeless individuals on the map.
- Providing analytical tools for the policymakers.
- Providing homeless growth predictions based on the available data.
- Performing shelter capacity utilization analysis, which can be used by policymakers to justify opening a new housing site or to close an existing site.
- Scheduling application to assist the homeless individual in future planning.
- Swapping, or moving homeless individuals from one housing provider to another to increase the overall Goodness-of-fit.

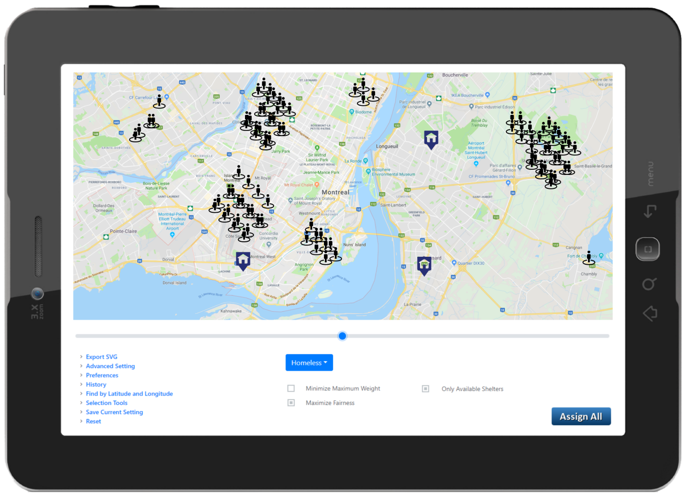

Figure 5.7: Policymaker application

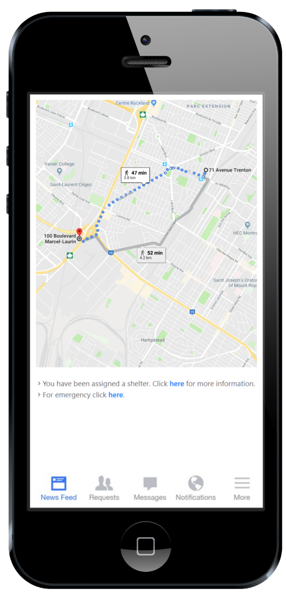

Figure 5.8: Client/homeless individual application

Chapter 6

Conclusion

In this thesis, we conducted two usability testing surveys for an AI-supported web-based simulation and modeling software designed to use computational intelligence to benefit members of the social sector, namely researchers and policy analysts. Additionally, we presented a computational model which could be used to alleviate some of the problems experienced by homeless individuals in both public and private temporary housing.

The first theme of this thesis discussed detailed steps of the design and development of a UI that was carefully planned over the course of 2 years in collaboration with three professors from multidisciplinary fields. A usability survey then was conducted with 14 participants to test the UI and user experience of the HOMVIZ application (Chapter 3). To the best of our knowledge, the HOMVIZ platform was one of the first in its class to use an aesthetically appealing UI element to help policymakers and researchers utilize AI algorithms. Our study focused on interface design, system navigation, and functionality and provided a detailed list of incorporated elements and practices that were used. The decision to conduct a survey with participants from multidisciplinary fields proved to be helpful as we noticed technical diversity led to rich and valuable feedback regarding various aspects of our software.

During our research, we learned that moderated sessions are difficult to plan and conduct. Therefore, we dedicated the next chapter (Chapter 4) of this thesis to compare the data quality collected from similar participant samples and for the same software during different usability testing sessions. In chapter 4, we compared the quality of data gathered in remote moderated and remote unmoderated usability testing sessions. This study is important to researchers as unmoderated sessions have various advantages over moderated sessions. Our study showed that unmoderated sessions had a larger occurrence of attrition and CR, however, because these issues were somewhat anticipated we programmed built-in functionalities to detect such events. We concluded that unmoderated sessions could be as effective if such measures are in place to identify and discard unreliable data.

The final theme of this research was an additional attempt to apply computational intelligence to problems faced in the social sector (Chapter 5). This study is important to researchers and policymakers as it introduces effective approaches that can be used to enhance the quality of life of individuals experiencing homelessness. This research provided eight heuristic algorithms that reduced the computational time by approximately 38 hours. Furthermore, several important factors such as goodness-of-fit and fairness were introduced which can be of great use to researchers within the computer science field.

The subject matter studied in this research was motivated by the desire to improve the collaboration between the computer science and the social sciences. This is an important connection to make as the advances made in the field of computer science can be applied to help confront the exceedingly complicated issues faced within the social sector. The proposed research in each chapter of this thesis can benefit the collaboration between these two fields, and in turn, help identify, address, and mitigate some of the problems that are faced in the social sector.

Bibliography

[1] Taressa Fraze et al. "Housing, transportation, and food: how ACOs seek to improve population health by addressing nonmedical needs of patients". In: *Health Affairs* 35.11 (2016), pp. 2109–2115.

[2] Emilio Serrano and Javier Bajo. "Deep neural network architectures for social services diagnosis in smart cities". In: *Future Generation Computer Systems* 100 (2019), pp. 122–131.

[3] Flaminio Squazzoni, Wander Jager, and Bruce Edmonds. "Social simulation in the social sciences: A brief overview". In: *Social Science Computer Review* 32.3 (2014), pp. 279–294.

[4] Vijay K Mago et al. "Analyzing the impact of social factors on homelessness: a Fuzzy Cognitive Map approach". In: *BMC medical informatics and decision making* 13.1 (2013), p. 94.

[5] Evan Arthur Ashkin. "Vulnerable Populations". In: *Chronic Illness Care*. Springer, 2018, pp. 331–341.

[6] Bernd W Wirtz, Jan C Weyerer, and Franziska T Schichtel. "An integrative public IoT framework for smart government". In: *Government Information Quarterly* 36.2 (2019), pp. 333–345.

[7] Riccardo Petrolo, Valeria Loscri, and Nathalie Mitton. “Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms”. In: *Transactions on Emerging Telecommunications Technologies* 28.1 (2017), e2931.

[8] Soledad Pellicer et al. “A global perspective of smart cities: A survey”. In: *2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*. IEEE. 2013, pp. 439–444.

[9] Muyiwa Lawrence Akinluyi and Adeniyi Adedokun. “Urbanization, Environment and Homelessness in the Developing world: The Sustainable Housing Development.” In: *Mediterranean Journal of Social Sciences* 5.2 (2014), pp. 261–261.

[10] Rita Yi Man Li and Jennifer Li. “The impact of subprime financial crisis on Canada and United States housing market and economy”. In: *ICBMG Conference*. Vol. 59. 2013, pp. 1–6.

[11] Patrick J Fowler et al. “Solving homelessness from a complex systems perspective: insights for prevention responses”. In: *Annual review of public health* 40 (2019), pp. 465–486.

[12] S Gaetz, T Gulliver, and T Richter. *The state of homelessness in Canada 2014: Canadian homelessness research network*. 2014.

[13] Eric A Latimer et al. “Costs of services for homeless people with mental illness in 5 Canadian cities: a large prospective follow-up study”. In: *CMAJ open* 5.3 (2017), E576.

[14] Julio Aráuz. “Smart Cities and the Dire Need for a Course Correction”. In: *2018 IEEE International Smart Cities Conference (ISC2)*. IEEE. Kansas City, MO, USA, Sept. 2018, pp. 1–6.

[15] Muyiwa Lawrence Akinluyi and Adeniyi Adedokun. "Urbanization, Environment and Homelessness in the Developing world: The Sustainable Housing Development." In: *Mediterranean Journal of Social Sciences* 5.2 (2014), p. 261.

[16] Rita Yi Man Li and Jennifer Li. "The impact of subprime financial crisis on Canada and United States housing market and economy". In: *ICBMG Conference*. Vol. 59. Hong Kong, 2013, pp. 1–6.

[17] Tanya Gulliver Garcia Stephen Gaetz and Tim Richter. *The state of homelessness in Canada 2014: Canadian homelessness research network*. Tech. rep. 978-1-77221-001-9. The Homeless Hub Press, Jan. 2014.

[18] Jill S. Roncarati et al. "Mortality Among Unsheltered Homeless Adults in Boston, Massachusetts, 2000-2009". In: *JAMA Internal Medicine* 178.9 (Sept. 2018), pp. 1242–1248. ISSN: 2168-6106. DOI: `10.1001/jamainternmed.2018.2924`. eprint: `https://jamanetwork.com/journals/jamainternalmedicine/articlepdf/2687991/jamainternal\_roncarati\_2018\_oi\_180046.pdf`. URL: `https://doi.org/10.1001/jamainternmed.2018.2924`.

[19] Eric Latimer and François Bordeleau. *Dénombrement des personnes en situation d'itinérance au Québec le 24 avril 2018*. Tech. rep. 18-846-10W. Montréal, QC: Ville de Montréal et Centre intégré universitaire de santé et de services sociaux du Centre-Sud-de-l'île-de-Montréal, Mar. 2019.

[20] Eric A Latimer et al. "Costs of services for homeless people with mental illness in 5 Canadian cities: a large prospective follow-up study". In: *CMAJ open* 5.3 (2017), E576.

[21] S Gaetz, T Gulliver, and T Richter. *The state of homelessness in Canada 2014: Canadian homelessness research network*. 2014.

[22] Peng Cui et al. “Uncovering and predicting human behaviors”. In: *IEEE Intelligent Systems* 31.2 (2016), pp. 77–88.

[23] Bill Edgar. “The ETHOS definition and classification of homelessness and housing exclusion”. In: *European Journal of Homelessness* 6.2 (2012), pp. 219–225.

[24] Stephen Gaetz and Erin Dej. *A new direction: A framework for homelessness prevention*. Canadian Observatory on Homelessness Press Toronto, 2017.

[25] Deborah Padgett, Benjamin F Henwood, and Sam J Tsemberis. *Housing First: Ending homelessness, transforming systems, and changing lives*. Oxford University Press, USA, 2016.

[26] Kelly L Page and Mark D Uncles. “The complexity of surveying web participation”. In: *Journal of Business Research* 67.11 (2014), pp. 2356–2367.

[27] Kholoud Al-Qeisi et al. “Website design quality and usage behavior: Unified Theory of Acceptance and Use of Technology”. In: *Journal of Business Research* 67.11 (2014), pp. 2282–2290.

[28] Christos Katsanos, Nikolaos Tselios, and Nikolaos Avouris. “Evaluating website navigability: validation of a tool-based approach through two eye-tracking user studies”. In: *New review of Hypermedia and Multimedia* 16.1-2 (2010), pp. 195–214.

[29] Yan Chen et al. “Improving Crowd-Supported GUI Testing with Structural Guidance”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–13.

[30] James R Lewis. “The system usability scale: past, present, and future”. In: *International Journal of Human–Computer Interaction* 34.7 (2018), pp. 577–590.

[31] Konstantina Orfanou, Nikolaos Tselios, and Christos Katsanos. “Perceived usability evaluation of learning management systems: Empirical evaluation of the System Usability Scale”. In: *The International Review of Research in Open and Distributed Learning* 16.2 (2015), pp. 227–246.

[32] Sven Coppers et al. “Intellingo: An Intelligible Translation Environment”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13.

[33] Karola Marky et al. “What did I really vote for? On the usability of verifiable e-voting schemes”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13.

[34] Mariam Hassib et al. “HeartChat: Heart rate augmented mobile chat to support empathy and awareness”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 2239–2251.

[35] Nathaniel Hudson et al. “Investigating how online help and learning resources support children’s use of 3D design software”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–14.

[36] Economic Advisors. *The State of Homelessness in America*. 2019.

[37] FEANTSA. *Third Overview of Housing Exclusion in Europe 2018*. 2018.

[38] Volker Busch-Geertsema, Dennis Culhane, and Suzanne Fitzpatrick. “Developing a global framework for conceptualising and measuring homelessness”. In: *Habitat International* 55 (2016), pp. 124–132.

[39] Marybeth Shinn and Jill Khadduri. *In the Midst of Plenty: Homelessness and what to Do about it*. John Wiley & Sons, 2020.

[40] Jeffrey Turnbull, Wendy Muckle, and Christina Masters. "Homelessness and health". In: *Cmaj* 177.9 (2007), pp. 1065–1066.

[41] T Aubry et al. "A randomized controlled trial in five Canadian cities of the effectiveness of Housing First with assertive community treatment for persons with serious mental illness and a history of homelessness". In: *Psychiatric Services* 67.3 (2016), pp. 81–275.

[42] Erin J Stringfellow et al. "Substance use among persons with homeless experience in primary care". In: *Substance abuse* 37.4 (2016), pp. 534–541.

[43] Philippe J Giabbanelli and Vijay K Mago. "Teaching computational modeling in the data science era". In: *Procedia Computer Science* 80 (2016), pp. 1968–1977.

[44] Sameera Abar et al. "Agent Based Modelling and Simulation tools: A review of the state-of-art software". In: *Computer Science Review* 24 (2017), pp. 13–33.

[45] Stewart Robinson. "A tutorial on simulation conceptual modeling". In: *2017 Winter Simulation Conference (WSC)*. IEEE. 2017, pp. 565–579.

[46] Jesús M Sánchez, Carlos A Iglesias, and J Fernando Sánchez-Rada. "Soil: an agent-based social simulator in Python for modelling and simulation of social networks". In: *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer. 2017, pp. 234–245.

[47] Seoyoung Hong and Jinwoo Kim. "Architectural criteria for website evaluation–conceptual framework and empirical validation". In: *Behaviour & Information Technology* 23.5 (2004), pp. 337–357.

[48] Suzanne Kieffer and Jean Vanderdonckt. "STRATUS: A questionnaire for strategic usability assessment". In: *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. 2016, pp. 205–212.

[49] James R Lewis. "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use". In: *International Journal of Human-Computer Interaction* 7.1 (1995), pp. 57–78.

[50] Nouzha Harrati et al. "Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis". In: *Computers in Human Behavior* 61 (2016), pp. 463–471.

[51] Andrew Fisher, Vijay Mago, and Eric Latimer. "Simulating the Evolution of Homeless Populations in Canada Using Modified Deep Q-Learning (MDQL) and Modified Neural Fitted Q-Iteration (MNFQ) Algorithms". In: *IEEE Access* 8 (2020), pp. 92954–92968.

[52] Vicky Stergiopoulos et al. "Effect of scattered-site housing using rent supplements and intensive case management on housing stability among homeless adults with mental illness: a randomized trial". In: *Jama* 313.9 (2015), pp. 905–915.

[53] Margaret G Hudson et al. *Interactive tooltip*. US Patent 6,828,988. Dec. 2004.

[54] Jeffrey M Rhodes. "Adding jQuery UI Controls". In: *Creating Business Applications with Office 365*. Springer, 2019, pp. 21–28.

[55] Martin Bean. *Laravel 5 essentials*. Packt Publishing Ltd, 2015.

[56] Jake Spurlock. *Bootstrap: Responsive Web Development*. " O'Reilly Media, Inc.", 2013.

[57] Eric Sarrion. *jQuery UI: Learn How to use Dialogs, Autocomplete, and More.* " O'Reilly Media, Inc.", 2012.

[58] Nick Qi Zhu. *Data visualization with D3. js cookbook*. Packt Publishing Ltd, 2013.

[59] Chadia Abras et al. "Encyclopedia of human-computer interaction". In: *Thousand Oaks: Sage Publications* 37 (2004), pp. 445–456.

[60] Nava Tintarev and Judith Masthoff. "Effective explanations of recommendations: user-centered design". In: *Proceedings of the 2007 ACM conference on Recommender systems*. 2007, pp. 153–156.

[61] Leming Zhou, Dilhari DeAlmeida, and Bambang Parmanto. "Applying a User-Centered Approach to Building a Mobile Personal Health Record App: Development and Usability Study". In: *JMIR mHealth and uHealth* 7.7 (2019), e13194.

[62] Meinald T Thielsch and Gerrit Hirschfeld. "Facets of website content". In: *Human–Computer Interaction* 34.4 (2019), pp. 279–327.

[63] Ian Robertson and Philip Kortum. "An Investigation of Different Methodologies for Rating Product Satisfaction". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 63. 1. SAGE Publications Sage CA: Los Angeles, CA. 2019, pp. 1259–1263.

[64] Helder Da Rocha. *Learn Chart. js: Create interactive visualizations for the web with chart. js 2*. Packt Publishing Ltd, 2019.

[65] Plotly. *Interactive Web-Based Data Visualization Library*. `https://plotly.com/`. 2020.

[66] Limon Monte. *Responsive, customizable, accessible replacement for javascript's popup boxes*. `https://sweetalert2.github.io/`. 2020.

[67] Ethan Marcotte. *Responsive web design: A book apart n 4*. Editions Eyrolles, 2017.

[68] David R Thomas. “A general inductive approach for analyzing qualitative evaluation data”. In: *American journal of evaluation* 27.2 (2006), pp. 237–246.

[69] Juan Cruz Gardey et al. “UX-Painter: An Approach to Explore Interaction Fixes in the Browser”. In: *Proceedings of the ACM on Human-Computer Interaction* 4.EICS (2020), pp. 1–21.

[70] Gyda Bjornsdottir et al. “From paper to web: Mode equivalence of the ARHQ and NEO-FFI”. In: *Computers in Human Behavior* 41 (2014), pp. 384–392.

[71] Morten Hertzum, Pia Borlund, and Kristina B Kristoffersen. “What do thinking-aloud participants say? A comparison of moderated and unmoderated usability sessions”. In: *International Journal of Human-Computer Interaction* 31.9 (2015), pp. 557–570.

[72] Aitor Apaolaza and Markel Vigo. “WevQuery: Testing hypotheses about web interaction patterns”. In: *Proceedings of the ACM on Human-Computer Interaction* 1.EICS (2017), pp. 1–17.

[73] Kapil Chalil Madathil and Joel S Greenstein. “An investigation of the efficacy of collaborative virtual reality systems for moderated remote usability testing”. In: *Applied ergonomics* 65 (2017), pp. 501–514.

[74] AJ Bernheim Brush, Morgan Ames, and Janet Davis. “A comparison of synchronous remote and local usability studies for an expert interface”. In: *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. 2004, pp. 1179–1182.

[75] XiaoChi Zhang et al. “Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ”. In: *Computers in Human Behavior* 71 (2017), pp. 172–180.

[76] Anja Mohorko and Valentina Hlebec. "Degree of cognitive interviewer involvement in questionnaire pretesting on trending survey modes". In: *Computers in Human Behavior* 62 (2016), pp. 79–89.

[77] MK Ward et al. "Careless response and attrition as sources of bias in online survey assessments of personality traits and performance". In: *Computers in Human Behavior* 76 (2017), pp. 417–430.

[78] Morten Hertzum, Rolf Molich, and Niels Ebbe Jacobsen. "What you get is what you see: revisiting the evaluator effect in usability tests". In: *Behaviour & Information Technology* 33.2 (2014), pp. 144–162.

[79] James R Lewis. *Introduction to the special issue on usability and user experience: methodological evolution*. 2015.

[80] Swati Sachan et al. "An explainable AI decision-support-system to automate loan underwriting". In: *Expert Systems with Applications* 144 (2020), p. 113100.

[81] Ahmed Kharrufa et al. "Group Spinner: recognizing and visualizing learning in the classroom for reflection, communication, and planning". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 5556–5567.

[82] Niels Ebbe Jacobsen, Morten Hertzum, and Bonnie E John. "The evaluator effect in usability studies: Problem detection and severity judgments". In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 42. 19. SAGE Publications Sage CA: Los Angeles, CA. 1998, pp. 1336–1340.

[83] Sharon McDonald, Gilbert Cockton, and Alastair Irons. "The impact of Thinking-Aloud on usability inspection". In: *Proceedings of the ACM on Human-Computer Interaction* 4.EICS (2020), pp. 1–22.

[84] Helen L Ball. "Conducting online surveys". In: *Journal of human lactation* 35.3 (2019), pp. 413–417.

[85] Hsiu-Mei Huang. "Do print and Web surveys provide the same results?" In: *Computers in Human Behavior* 22.3 (2006), pp. 334–350.

[86] Chia-Chi Wang et al. "Comparison of web-based versus paper-and-pencil administration of a humor survey". In: *Computers in Human Behavior* 29.3 (2013), pp. 1007–1011.

[87] Alexander S McKay et al. "The attentive and the careless: Examining the relationship between benevolent and malevolent personality traits with careless responding in online surveys". In: *Computers in Human Behavior* 84 (2018), pp. 295–303.

[88] Mary Kathrine Ward and Samuel B Pond III. "Using virtual presence and survey instructions to minimize careless responding on Internet-based surveys". In: *Computers in Human Behavior* 48 (2015), pp. 554–568.

[89] Andraž Petrovčič, Gregor Petrič, and Katja Lozar Manfreda. "The effect of email invitation elements on response rate in a web survey within an online community". In: *Computers in Human Behavior* 56 (2016), pp. 320–329.

[90] Tristan W Casey and Arthur Poropat. "Beauty is more than screen deep: Improving the web survey respondent experience through socially-present and aesthetically-pleasing user interfaces". In: *Computers in Human Behavior* 30 (2014), pp. 153–163.

[91] Weimiao Fan and Zheng Yan. "Factors affecting response rates of the web survey: A systematic review". In: *Computers in human behavior* 26.2 (2010), pp. 132–139.

[92] Mark Micallef, Chris Porter, and Andrea Borg. "Do exploratory testers need formal training? An investigation using HCI techniques". In: *2016 IEEE Ninth International*

Conference on Software Testing, Verification and Validation Workshops (ICSTW). IEEE. 2016, pp. 305–314.

[93] James R Lewis, Joshua Brown, and Daniel K Mayes. “Psychometric evaluation of the EMO and the SUS in the context of a large-sample unmoderated usability study”. In: *International Journal of Human-Computer Interaction* 31.8 (2015), pp. 545–553.

[94] Ulrich Schroeders and Oliver Wilhelm. “Computer usage questionnaire: Structure, correlates, and gender differences”. In: *Computers in Human Behavior* 27.2 (2011), pp. 899–904.

[95] Pedram Khayyatkhoshnevis et al. “Smart City Response to Homelessness”. In: *IEEE Access* 8 (2020), pp. 11380–11392.

[96] John Travis Marshall and Jessica Venegas. “The Smart Cities Movement and Advancing the International Battle to Eliminate Homelessness-Barcelona as Test Case”. In: *Revista de Derecho Urbanıéstico y Medio Ambiente* (2015).

[97] Myra Piat et al. “Pathways into homelessness: Understanding how both individual and structural factors contribute to and sustain homelessness in Canada”. In: *Urban Studies* 52.13 (2015), pp. 2366–2382.

[98] Jon Kleinberg and Eva Tardos. *Algorithm design.* Pearson Education India, 2006.

[99] Qiuping Li et al. “Energy-efficient computation offloading and resource allocation in fog computing for internet of everything”. In: *China Communications* 16.3 (2019), pp. 32–41.

[100] Arent W de Jong et al. “Big data in automation: Towards generalized makespan estimation in shop scheduling problems”. In: *2017 13th IEEE Conference on Automation Science and Engineering (CASE).* IEEE. Xi’an, China, 2017, pp. 1516–1521.

[101] Seyed Mohammad Hossein Hojjati and Amin Sahraeyan. “Minimizing makespan subject to budget limitation in hybrid flow shop”. In: *2009 International Conference on Computers & Industrial Engineering*. IEEE. Troyes, France, 2009, pp. 18–22.

[102] Poonam Singh, Maitreyee Dutta, and Naveen Aggarwal. “Bi-objective HWDO Algorithm for Optimizing Makespan and Reliability of Workflow Scheduling in Cloud Systems”. In: *2017 14th IEEE India Council International Conference (INDICON)*. IEEE. Roorkee, India, 2017, pp. 1–9.

[103] Majid M Khodier and Christos G Christodoulou. “Linear array geometry synthesis with minimum sidelobe level and null control using particle swarm optimization”. In: *IEEE Transactions on Antennas and propagation* 53.8 (2005), pp. 2674–2679.

[104] Ayoub Alami Masmoudi and Mohamed Benbrahim. “New heuristics to minimize makespan for two identical parallel machines with one constraint of unavailability on each machine”. In: *2015 International Conference on Industrial Engineering and Systems Management (IESM)*. IEEE. Seville, Spain, 2015, pp. 476–480.

[105] Jingying Huang and Liya Wang. “Makespan minimization on single batch-processing machine considering preventive maintenance”. In: *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*. IEEE. Singapore, 2018, pp. 294–298.

[106] Arent W De Jong et al. “A generalised makespan estimation for shop scheduling problems, using visual data and a convolutional neural network”. In: *International Journal of Computer Integrated Manufacturing* (2019), pp. 1–10.

[107] Dikscha Sapra, Rashi Sharma, and Arun Prakash Agarwal. “Comparative study of metaheuristic algorithms using Knapsack Problem”. In: *2017 7th International Con-*

ference on Cloud Computing, Data Science & Engineering-Confluence. IEEE. Noida, India, 2017, pp. 134–137.

[108] *IBM: ILOG CPLEX Optimization Studio 12.7.1*. 2014. URL: `http://ibm.biz/COS1271Documentation`.

[109] Rajendra K Jain, Dah-Ming W Chiu, and William R Hawe. “A quantitative measure of fairness and discrimination”. In: *Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA* (1984).

Appendix A

Questionnaires

Table A.1: The System Usability Scale Questionnaire (SUS) [30].

	The System Usability Scale Questionnaire. Please check the boxes that reflect your immediate response to each statement.	**Strongly Disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
(1)	I think that I would like to use this web application frequently.	☐	☐	☐	☐	☐
(2)	I found the web application unnecessarily complex.	☐	☐	☐	☐	☐
(3)	I thought the web application was easy to use.	☐	☐	☐	☐	☐
(4)	I think that I would need the support of a technical person to be able to use this web application.	☐	☐	☐	☐	☐
(5)	I found the various functions in this web application were well integrated.	☐	☐	☐	☐	☐
(6)	I thought there was too much inconsistency in this web application.	☐	☐	☐	☐	☐
(7)	I would imagine that most people would learn to use this web application very quickly.	☐	☐	☐	☐	☐
(8)	I found the web application very cumbersome/awkward to use.	☐	☐	☐	☐	☐
(9)	I felt very confident using the web application.	☐	☐	☐	☐	☐
(10)	I needed to learn a lot of things before I could get going with this system.	☐	☐	☐	☐	☐
(11)	What are some aspects of the HOMVIZ platform which you found to be positive?	☐	☐	☐	☐	☐
(12)	What are some aspects of the HOMVIZ platform which you found to be negative?	☐	☐	☐	☐	☐

Table A.2: The Computer Usage Questionnaire (CUQ) [94].

		Never	**Rarely**	**Some times**	**Often**	**Very often**
(1)	**How often do you use the following programs?**					
(a)	Word processing (e.g., Word)	☐	☐	☐	☐	☐
(b)	Spreadsheet (e.g., Excel)	☐	☐	☐	☐	☐
(c)	Presentation program (e.g., Powerpoint)	☐	☐	☐	☐	☐
(d)	Programming language (e.g., Java)	☐	☐	☐	☐	☐
(e)	Graphics software (e.g., CorelDraw)	☐	☐	☐	☐	☐
(f)	Sound or video editing software	☐	☐	☐	☐	☐
(g)	e-mail client (e.g., Outlook)	☐	☐	☐	☐	☐
(h)	Chat program (e.g., IRC, Skype)	☐	☐	☐	☐	☐
(i)	Web browser (e.g., Firefox, IE)	☐	☐	☐	☐	☐
(j)	Games (e.g., The Sims)	☐	☐	☐	☐	☐
(2)	**How often do you perform the following computer activities?**					
(a)	Creating a presentation	☐	☐	☐	☐	☐
(b)	Programming	☐	☐	☐	☐	☐
(c)	Sound editing	☐	☐	☐	☐	☐
(d)	Writing e-mails	☐	☐	☐	☐	☐
(e)	Chatting	☐	☐	☐	☐	☐
(f)	Surfing the web	☐	☐	☐	☐	☐
(g)	Playing alone	☐	☐	☐	☐	☐
(h)	Playing online	☐	☐	☐	☐	☐

Appendix B

Research Ethics Application

Figure B.1: Research ethics application form

ROMEO - Researcher Portal

Research Ethics Board Researcher's Agreement Form

Project Info.

File No: 1468374
Project Title: Evaluation of moderated vs unmoderated software usability study
Principal Investigator: Mr. Pedram khayyatKhoshnevis (Science and Environmental Studies\Computer Science)
Start Date: 2020/12/01
End Date: 2021/11/01
Keywords: software usability, simulation, social systems, moderated survey, unmoderated survey

Project Team Info.

Principal Investigator

Prefix: Mr.
Last Name: khayyatKhoshnevis
First Name: Pedram
Affiliation: Science and Environmental Studies\Computer Science
Position: RES
Email: pkhayyat@lakeheadu.ca
Phone1: 3065803888
Phone2:
Fax:
Primary Address: 136 Sherbrooke Street
Institution: Lakehead University
Country: Canada
Comments: Additional project team members:Ms. Savanah Tillberg | Research Assistant | Lakehead University

Other Project Team Members

Prefix	Last Name	First Name	Affiliation	Role In Project	Email
Dr.	Latimer	Eric	External\External	Co-Investigator	eric.latimer@douglas.mcgill.ca

Dr.	Mago	Vijay	Science and Environmental Studies\Computer Science	Supervisor	vmago@lakeheadu.ca

Common Questions

1. General Information

#	Question	Answer
1.1	Type of Participants:	Adults
1.2	Estimated Number of Participants (#):	100
1.3	Where will the research be conducted?	Lakehead Campuses (Thunder Bay or Orillia)
1.4	Have you received approval, or are you seeking approval from any other ethics committee?	No
1.5	If "Yes", which Ethics Committee(s)?	
1.6	Is this project's funding administered through Lakehead University?	Yes
1.7	Is this project's funding administered outside of Lakehead University?	No
1.8	If funded, Name of Granting Agency:	SSHRC Insight grant

2. Preliminary Checklist

#	Question	Answer
2.1	Will your study involve more than minimal physical risk to your participants? (For a definition of minimal risk, see TCPS2, Chapter 2, Section B.)	No
2.2	Will your study involve the use of high-risk test instruments, i.e. surveys that may reveal that the participant intends to participate in dangerous activities such as self harm or harm to others?	No
2.3	Will your study involve more than minimal psychological risk to your participants?	No
2.4	Will your study likely lead to the discovery of your participants' involvement in illegal activities?	No
2.5	Will your study involve participants who are members of vulnerable populations?	No
2.6	If "Yes" to any of the above, please elaborate briefly:	

2.7	Will your study involve clinical research?	No
2.8	If "Yes", please elaborate briefly:	
2.9	Will your study involve First Nation, Metis and/or Inuit (FNMI) peoples as a "scheduled" representative group in the research?	No
2.10	If "Yes", describe your collaboration or community engagement plan for guiding and monitoring the research with appropriate Aboriginal (FNMI) groups, including how the community has been engaged and plans for future engagement. For FNMI people residing outside of First Nations communities (ie. in urban centres), a similar plan of engagement with a representative FNMI group is required.	
2.11	Will your study involve, effect or impact Aboriginal (FNMI) peoples as an "incidental" representative group in research?	No
2.12	If "Yes", please refer to TCPS 2, Article 9.2 and indicate your awareness of how your study may involve or impact FNMI peoples, in the short or long term. Demonstrate awareness by describing how your study will involve, effect or impact FNMI peoples. For example, is there a benefit of FNMI peoples in the longer-term research outcomes?	
2.13	I confirm that if I become aware that my participant pool contains a sizeable proportion of FNMI participants, I will advise the REB of my new research situation by submitting an amendment/addition through the Romeo Research Portal.	Yes

3. Research Ethics Review Criteria

#	Question	Answer

3.1	LAY DESCRIPTION: Provide a brief lay-word summary of the proposed project (40 words or less, similar to the statement you would prepare for a granting agency for public dissemination).	The proposed project will evaluate and compare the web software usability feedback and questionnaire results for moderated and unmoderated survey methods. All sessions are conducted online using the Zoom application.

3.2	SUMMARY OF PURPOSE OF RESEARCH: Be sure to include sufficient detail, described in terms that do not require extensive field-specific knowledge. Include your research question(s).	A user study is a vital part of the design and development of an effective user interface (UI) and rich user experience (UX). Numerous studies suggest that there is a difference in data quality depending on the type of survey used, namely online or offline. For example, various research suggests that surveys that are conducted using the online method suffer from a phenomenon called the careless response. Similarly, many previous studies have been conducted to compare unmoderated versus moderated surveys. However, those studies addressed the usability of simple-to-use web platforms such as e-commerce, and news websites.Based on our findings, there is a lack of research regarding the correlation of survey methods and data quality for complex software such as web simulation and modeling tools. A simulation and modeling tool is software that models real-life phenomena, like homelessness, using mathematical relationships. Models allow users to observe the phenomenon in a simulation without it happening in real life. This type of software is used in many fields, including healthcare and government, and it enables researchers and policymakers to study the effects of certain events and/or policy decisions.Simulation and modeling software differs from other types of web platforms as it requires a pre-training stage where the users learn how to use the tool. During this stage potential users increase their confidence in their understanding of the different variables and parameters required to create a model and run a simulation. Given the complex nature of such software, most usability sessions for simulation and modeling tools are conducted with the help of a moderator. The moderator introduces the system and

		guides users through a tutorial session where they closely monitor users' behaviors and answer their questions during each session. It is apparent that a usability study for simulation and modeling software is best conducted when moderated by a professional with the knowledge necessary to aid users throughout each step. Because unmoderated surveys are more cost-effective and less time-consuming, research that investigates whether simulation and modeling tools can benefit from online and unmoderated survey methods is needed.To find answers to this question we will conduct a randomized controlled trial testing the effectiveness and level of reliability of the data collected of two survey modes, namely moderated and unmoderated.
3.3	RESEARCH PARTICIPANTS: Describe required characteristics and number of participants.	We are recruiting approximately 100 participants from the department of computer science at Lakehead University who are familiar enough with computers to comfortably use our website. The participants will be equally divided into two random groups using a block randomization method.

3.4	DATA COLLECTION: Explain the method of data collection and analysis. Explain exactly what will be expected of participants (length of time commitment, etc.) All questionnaires and research instruments must be included as appendices.	We are using various methods to compare the results of moderated and unmoderated surveys. During both sessions, participants are asked to sign up to our website using their email addresses. On the registration page, we also collect participants' age and gender (male, female, other, prefer not to answer). Age and gender are used to describe the characteristics of survey participants. We plan to collect various types of data from each usability session. This includes the mouse movement of users on the task's page only, the time it takes to create a simulation model, the time they spend watching the tutorial video (for unmoderated study only), the time they spend answering the questionnaires, and the questionnaire results.Participants are asked to complete two questionnaires; the System usability questionnaire (SUS) and the Computer usage questionnaire (CUQ). The SUS questionnaire will question participants regarding their experience with our system. The SUS mainly focuses on user satisfaction and the user-friendliness of our platform. The CUQ questionnaire asks participants about their computer skills and experience, for example, how often they use computers on a daily basis.The mouse movement is analyzed to study the users' confidence level in accomplishing the task among moderated and unmoderated sessions. The completion time of various parts of the study is used to compare the effort participants made in each survey method; mainly studying the careless response effect in the unmoderated study. The questionnaires are also compared to find a further correlation between the type of survey and user confidence in their ability and satisfaction with our platform.

3.5	SECONDARY DATA: For research involving the use of secondary data (data which has been previously collected for a purpose other than the research project itself), REB review is not required if the data is anonymized so long as the process of data linkage or recording or dissemination of results does not generate identifiable information (see TCPS 2, Chapter 2, Article 2.4). For secondary data that is identifiable, please see TCPS 2, Chapter 5, Section D.	No secondary data will be used.
3.6	RECRUITMENT PROCEDURES: Describe how potential participants will be selected and contacted. Include a copy of any advertisements used to recruit participants.	The participants will be recruited from a pool of undergraduate and graduate students at Lakehead University.Instructors of courses COMP 5111 Graduate Seminar, COMP 4431 Big Data, COMP 4431/32/34 - Advanced Project, and COMP 4475 Topics in AI will be contacted to allow us to conduct the surveys.
3.7	HARM and/or POTENTIAL RISKS to PARTICIPANTS: (a) State clearly any potential harm or risks - physical, psychological, injury to reputation or privacy, and breach of any relevant law - for participants or for third parties (those affected by the research but who are not active research subjects); (b) If there is any apparent or potential harm or risk, clearly explain all steps that are being taken to reduce this.	There are no expected physical or psychological risks to participants.
3.8	DECEPTION: If deception is part of the research program, the researcher must: (a) State clearly why no alternative methodology, which does not involve deception, can fruitfully be used to answer the research question; (b) Provide evidence that the participant is not put at risk by the deception. If appropriate, provide a debriefing letter to participants disclosing the deception.	Deception is not a part of this research.

3.9	BENEFITS to PARTICIPANTS and/or SOCIETY: Describe in detail the potential benefits of the research for both participants and to general knowledge.	The potential benefit of this study is a means of providing researchers in the same field the comparison results of the data quality gathered in moderated and unmoderated usability surveys. Unmoderated studies offer the advantage of being time and cost-effective. However, their effectiveness for social simulation and modeling software remains unknown.Participants of these surveys are recruited from undergraduate and graduate courses in the computer science department. The learning outcomes of these courses align with the nature of this project. Students will gain relevant experiential learning by participating in this project. In compensation for students' time, using a random draw, three participants will be selected for a $20 Amazon gift card each.
3.10	INFORMED CONSENT: Clearly outline the measures that will be used to ensure the informed consent of all research participants. Cover letters and consent forms must be attached as appendices on Lakehead University (or NOSM if appropriate) letterhead.	Before beginning the survey, participants will be required to confirm they have read the information letter and consent letter by clicking a box confirming their understanding and willingness to consent to participate in this research.
3.11	CAPACITY TO CONSENT: Capacity refers to the ability of prospective or actual participants to understand relevant information presented about a research project, and to appreciate the potential consequences of their decision to participate or not participate (TCPS 2, Chapter 3, Section C). Will the research participants sufficiently understand the nature of the research project, and the risks, consequences, and potential benefits associated with it?	Yes
3.12	If "No", please state why this vulnerable group is necessary to the study, and elaborate on the consent process (i.e. parental consent, caregiver consent).	

3.13	RIGHT TO WITHDRAW: The researcher must illustrate that participants will be informed of their right to withdraw from the study at any time without penalty of any kind, and that they may choose not to answer any question asked as part of the research. For participants submitting information anonymously, the participants must be informed that withdrawal post-submission is not possible due to the anonymous nature of their data. Will the participants have the right to withdraw?	Yes
3.14	If "No", please elaborate:	
3.15	ANONYMITY and/or CONFIDENTIALITY: The researcher must outline the procedures that will be used to guarantee confidentiality and/or anonymity for participants. Participants who wish to be named and to waive their right to privacy and confidentiality must provide written evidence.	Participants' emails will be stored separately and securely on a password-protected database. The raw survey results will not be shared publicly and it is the only used to find a correlation between moderated and unmoderated surveyed.
3.16	STORAGE of DATA: Detail how the data will be securely stored for a minimum of 5 years following completion of the research, at Lakehead University, as per Lakehead University policy.	The email addresses will be stored separately on an encrypted and password-protected database. Other collected information such as survey response data will be stored on secure password-protected servers of DaTALab Lakehead University for a period of 5 years under the responsibility of Dr. Vijay Mago, the supervisor for this study.
3.17	PEER REVIEW: State the intention, or non-intention to have the proposal peer reviewed by an external granting agency or thesis committee, if appropriate. If the REB determines the project to be of more than minimal risk, peer review may be required by the addition of ad-hoc members to the REB, even if the granting agency for the project does not require this, or if the project is not funded.	We do not seek additional peer review for this project.

3.18	RESEARCH PARTNERS and STUDENTS: Clearly state whether or not the research will involve student researchers or if the research will involve researchers at another university/institution. TCPS2 Tutorial Certificates for all research partners and students must be attached.	Student researchers and research partners from other universities (Mcgill, and Ottawa University) will be involved. We have provided their TCPS2 certificates.
3.19	MULTI-JURISDICTION RESEARCH: If you are involved in multi-jurisdictional research, provide evidence that ethical approval is also being sought at any other institution where direct research with human participants will be undertaken. Ethical approval from another institution, while essential in a multi-jurisdiction project, is not itself sufficient for the commencement of research with human participants at Lakehead University.	Our study does not involve multi-jurisdictional research.
3.20	CONFLICT of INTEREST: Disclose any real, perceived or potential conflicts of interest (professional, personal or financial) to the Research Ethics Board. NOTE: It is preferable to avoid or prevent being in a conflict of interest, when possible. When it is not possible to avoid a conflict of interest, then it should be disclosed to the appropriate people and steps taken to minimize or manage the conflict.	We have no conflict of interest to declare.
3.21	DISSEMINATION of RESEARCH RESULTS: Clearly state the means by which the research will be disseminated and by which research participants may be made aware of the findings of the study. When requested by participants, research results must be provided in easily understandable language and approved by the Principal Investigator.	Research results will be disseminated in publications in professional conferences and journals. Open-access journals will be favored, thus providing access to our findings to all participants.
3.22	I have completed the TCPS 2 Tutorial: Course on Research Ethics (CORE) and have attached a copy of my completion certificate to this application. *Please note that all investigators listed on this application must submit their certificates.	Yes

3.23	I am familiar with the Agreement on the Administration of Agency Grants and Awards by Research Institutions, and the Tri-Council Policy Statement 2: Ethical Conduct for Research Involving Humans and I agree to comply with these guidelines, and the procedures approved by the REB, in carrying out this proposed research.	Yes
3.24	I attest that all information submitted to the REB is complete and truthful. I understand the consequences, for myself and for the institution, of failure to comply with the above regulations.	Yes
3.25	Researchers are required to report to the REB any changes in research design, procedures, sample characteristics, and so forth that are contemplated after REB approval has been granted. Changes may not be implemented until approved by the REB. If any unforeseen incident occurs during the course of research that may indicate risk to participants, I will immediately cease research and inform the REB.	Yes
3.26	I understand that my protocol will be subject to random review for compliance by the Office of Research Services.	Yes
3.27	I will inform the REB when the research is complete by completing the Final Report Form (see New Event Forms in the Romeo Research Portal).	Yes

4. COVID-19 Measures

#	Question	Answer
4.1	Temporarily Unavailable	

5. Informed Consent Checklist (to assist Applicants)

#	Question	Answer

5.1	General	Cover letters and consent forms are presented on Lakehead University letterhead (or NOSM if appropriate)|The language level is appropriate to the age and reading level of the participant population|Contact information for the researcher(s) (including the supervisor if a graduate student project) and the REB is always included in the cover letter that the participants will keep after they sign the consent form. Suggested wording: "This study has been approved by the Lakehead University Research Ethics Board. If you have any questions related to the ethics of the research and would like to speak to someone outside of the research team please contact Sue Wright at the Research Ethics Board at 807-343-8283 or research@lakeheadu.ca."|If the PI on this application is NOT a department chair, NO action is required on the "Approvals" tab.

5.2	The Cover Letter/Introductory Information (including electronic letters and consent forms) should include:	The title of the study|An explanation of the purpose of the research|The identity of the researcher and their affiliation with Lakehead University|The funder of the research, if applicable|A warm, non-coercive invitation to participate, addressed to the "Potential Participant"|The reason why the potential participant is being invited to participate in the research|That the individual's participation is voluntary, that they may refuse to participate in any part of the study, and that they may withdraw from the study at any time (other than anonymously submitted information)|That participants may decline to answer any question|A description of the procedures the participants will be involved in and how much of their time will be required|Information regarding any audio or videotaping and explicit consent to such recording|Information about any foreseeable risks, harms, or inconveniences|Potential benefits (including information that there is no direct benefit, if appropriate)|A mechanism for providing referrals, if appropriate (ie. if there is the possibility of emotional distress or physical harm)|Information regarding who will have access to the data|Information about the storage of data (during and after completion of the research)|The degree of confidentiality and/or anonymity that will be provided and how this will be maintained (eg. individual participants will not be identified in published results without their explicit consent, data will be published in aggregate form). For research involving anonymous surveys, it should be stated that the survey instrument will not be labelled to identify who completed it |Limits on confidentiality, if applicable (eg. confidentiality disclaimer for focus

		groups)|A statement indicating the researcher's intent to publish or make presentations based on the research and whether or not the participant's identify will remain confidential (eg. will pseudonymous be used?)|Offer of summary of the results (and a mechanism to provide the summary)
5.3	The Consent Form must state each individual's agreement that:	They have read and understood the cover/information letter for the study|They agree to participate|They understand the potential risks and/or benefits of the study, and what those are|That they are a volunteer and can withdraw from the study at any time, and may choose not to answer any question|The data they provide will be securely stored at Lakehead University for a minimum of 5 years following completion of the project|If applicable, that they understand that the research findings will be made available to them, and how this will be communicated|That they will remain anonymous in any publication/public presentation of research findings. Participants must explicitly agree to have their identities revealed.

5.4	Other Consent Information	All participants must sign and date the consent form then return it to the researcher|Consent must also be obtained from all agencies, partners, schools, school boards, etc. that provide access to the subject pools. Separate consent forms must be included for all of the above should this apply|If the study involves the use of high-risk test instruments which could potentially reveal that the subject intends to participant in a dangerous activity(s), the consent should contain a clause such that there is a limit to the level of confidentiality when the subject may be at risk for harm to self or others|While inclusive research is important, the researcher must ensure that consent is obtained from vulnerable populations in a sensitive manner. Vulnerable populations include children, & others not competent to give free and informed consent on their own behalf. In cases like this, parent/guardian (or the individual's representative) consent must be obtained. Please note every effort should be made to ensure that participants understand and consent to their own participation as well. If representative consent will be obtained the researcher must explicitly demonstrate why this is necessary and how the research results could be significantly altered if representative consent was required.

Attachments

Doc / Agreement	**Version Date**	**File Name**	**Description**
Supporting Documents (REB)		Survey questions.pdf	Survey questions

Supporting Documents (REB)		Ethics form usability comparison EL signed (1) (1).pdf	signature page
Supporting Documents (REB)		Email invitation content.pdf	Email invitation template to invite participants to our survey. The file includes both moderated and unmoderated email templates.
Supporting Documents (REB)		pre-registartion_consent-form.PNG	Screenshot of the pre-registration page where participants must confirm reading and agree to the information letter and the consent form.
Supporting Documents (REB)		Unmoderated survey tasks.pdf	Task sheet that participants must follow to complete the unmoderated sruvey.
Supporting Documents (REB)		Moderated survey tasks.pdf	Task sheet that participants must follow to complete the moderated sruvey.
Supporting Documents (REB)		Infomation Letter.pdf	Information letter
Supporting Documents (REB)		Consent Form.pdf	Consent form
TCPS Certificate		tcps2_core_certificate_Pedram_Khoshnevis.pdf	Pedram Khayyatkhoshnevis's tcps2 core certificate
TCPS Certificate		tcps2-eptc2-certificate_Savanah_Tillberg.pdf	Savanah tillberg's tcps2 core certificate

TCPS Certificate		tcps2_core_certificate _Eric_Latimer.pdf	Eric latimer's tcps2 core certificate
TCPS Certificate		tcps2_core_certificate _Vijay_Mago.pdf	Vijay mago's tcps2 core certificate

Figure B.2: Survey information letter (page 1)

Department of Computer Science
t: (807) 343-8310
e: vmago@lakeheadu.ca

Participant Information Letter

Dear Potential Participant:

On behalf of researchers at Lakehead University, you are being invited to complete the following survey. This survey involves using simulation and modeling web software and filling up two questionnaires - System Usability Scale (SUS) and Computer Usage Questionnaire (CUQ). These questionnaires tell us about your perception of our system and also your computer skills. Completing the survey will approximately take 45 minutes to 1 hour of your time.

Taking part in this study is voluntary and you may withdraw at any time (only up until the point of submission of data), without penalty. Please note that a decision to participate or withdraw from this survey will have no effect on your academic status.

Before you decide whether or not you would like to take part in this study, please read this letter carefully to understand what is involved.

TITLE OF STUDY
Evaluation of moderated vs unmoderated software usability study

PURPOSE
A software usability survey is used to determine the user-friendliness and intuitiveness of software. There are two methods of usability study; moderated and unmoderated. In this research, we are interested in comparing the results of these two different methods and evaluating the data quality and accuracy of each survey method.

WHAT INFORMATION WILL BE COLLECTED?
We collect information from the questionnaire answer sheets, the results of the accomplished tasks, and the mouse movements only within the “create simulation page”. We also collect the email, gender, and age of our participants. Age and gender are used to describe the characteristics of survey participants.

WHAT IS REQUESTED OF ME AS A PARTICIPANT?
As previously mentioned, our survey compares moderated versus unmoderated usability sessions. You may be invited to participate in either of the survey sessions. In the moderated session, you will join us on a Zoom call (link will be sent at a later time)

Figure B.3: Survey information letter (page 2)

where we will present you with a quick live demo of how to use our web platform. You will then be given a task list that provides you with step-by-step instructions on how to create your own simulation model. Once you have completed your simulation, you will be redirected to a questionnaire page where you will be asked to answer several short questions. During the moderated session, we will be present to answer your questions. In the unmoderated session, you will receive an email containing a link to a longer task sheet. There, you will be provided with a link to a tutorial video on Youtube where you will learn how to use our platform. Following this, you will be asked to move on to the next step on the task sheet which provides you with step-by-step instruction on how to create your simulation model on our website. Following the completion of your simulation model, you will be asked to open the questionnaire page and answer the questions. We estimate that your participation in any of the modes will take 45 minutes to 1 hour of your time.

WHAT ARE THE RISKS AND BENEFITS?
There are no foreseeable risks or impacts on the academic status for participating in this survey. Benefits include standing a chance to win a $20 Amazon gift card and helping researchers answer questions regarding survey methods.

HOW WILL MY CONFIDENTIALITY BE MAINTAINED?
The data collected from the survey will be stored on secure password-protected servers of DaTALab at Lakehead University. Any results released from this study will only present aggregated data in such a way that no individual participant can be identified.

WHAT WILL MY DATA BE USED FOR:
The data will be used to compare two methods of surveys; namely moderated and unmoderated.

WHERE WILL MY DATA BE STORED?
The survey data will be stored on the secure servers, managed by Dr. Vijay Mago, for a period of 5 years.

HOW CAN I RECEIVE A COPY OF THE RESEARCH RESULTS?
The research results will be searchable under the works of Dr. Vijay Mago, once the research is complete and academic publications are prepared and submitted. Participants can also contact Dr. Vijay Mago at vmago@lakeheadu.ca to receive a direct copy of the results.

WHAT IF I WANT TO WITHDRAW FROM THE STUDY?
Taking part in this study is voluntary and you may withdraw at any time, without penalty, only up until the point of submission of data.

955 Oliver Road, Thunder Bay, ON, Canada, P7B 5E1 | lakeheadu.ca

Figure B.4: Survey information letter (page 3)

RESEARCHERS CONTACT INFORMATION:
Dr. Vijay Mago (vmago@lakeheadu.ca)
Dr. Eric Latimer (eric.latimer@mcgill.ca)
Mr. Pedram Khoshnevis (pkhayyat@lakeheadu.ca)
Ms. Savanah Tillberg (stillber@lakeheadu.ca)

There will be no commercialization of the research results. There are no conflicts of interest for the researchers involved.

RESEARCH ETHICS BOARD REVIEW AND APPROVAL:
This study has been approved by the Lakehead University Research Ethics Board. If you have any questions related to the ethics of the research and would like to speak to someone outside of the research team please contact Sue Wright at the Research Ethics Board at 807-343-8283 or research@lakeheadu.ca.

955 Oliver Road, Thunder Bay, ON, Canada, P7B 5E1 | lakeheadu.ca

Figure B.5: Survey consent form

Lakehead UNIVERSITY

Department of Computer Science
t: (807) 343-8310
e: vmago@lakeheadu.ca

Consent Form for HOMVIZ software usability Survey

Consent to Participate

By marking the consent checkbox on the webpage, I indicate that I have read the "Participant Information Letter" presented on this webpage and that I understand and agree to the following:

1. I understand the information contained in the "Participant Information Letter".
2. I agree to participate in the survey.
3. I understand the risks and benefits of this survey.
4. I understand that I am free to withdraw from the survey without losing my chance to win a $20 Amazon gift card.
5. I understand that the Information collected from this survey will be securely stored at Lakehead University, Thunder Bay, ON for a period, of maximum length, of 5 years following completion of the research project.
6. By clicking the box on the webpage I am indicating that I understand and agree to this "Consent to Participate".

Figure B.6: Invitation to moderated usability sessions (page 1)

Department of Computer Science
t: (807) 343-8310
e: vmago@lakeheadu.ca

Dear Potential Participant,

As a part of a study investigating different survey modes, we are conducting a web-based survey on a simulation platform. The purpose of this is to determine whether there is a difference in the quality of response and data for moderated and unmoderated usability testing. We are hopeful that this data will aid future researchers in their development of simulation software development.

On behalf of myself, and the rest of our team, I would like to invite you to participate in the web-based study. By taking part in this endeavour you will be contributing to future developments in simulation software algorithms and technology as well as standing a chance to win a $20 Amazon gift card.

We would greatly appreciate your honest feedback regarding your experience with our software so that we can make improvements to our system. It is important to mention that whether you join this survey or not your academic status or relationship with the Department and Lakehead University will not be affected. Additionally, since this is a volunteer-based survey there will be no academic credit for your participation.

The creation of the simulation scenario and a subsequent survey will take approximately one hour of your time. You will be invited to a meeting with our researchers via Zoom where you will be given a tutorial on how to use our software. Following this tutorial, while remaining on the Zoom call, you will be asked to attempt to create a simulation scenario using our website. The researchers will be present to aid you if you are to have any issues or questions. Following your completion of the simulation, you will be asked to complete a short usability questionnaire.

Prior to our meeting please read the information letter and consent form attached to this email. Later during our session, you will be asked if you read, understood, and agreed to the terms on the information letter and consent form. The information letter and consent form are also available on https://homviz.datalab.science/terms.

To join the Zoom meeting on [Date], please notify us by simply replying to this email.

Link to the Information Letter:
https://drive.google.com/file/d/1-1NnB6zcRGOI6GGeK7llobRvu2431WRx/view?usp=sharing

Link to the Consent Form:
https://drive.google.com/file/d/1q4n0W97z-TVX146cq9ij-lZZoUzV1Lde/view?usp=sharing

Figure B.7: Invitation to moderated usability sessions (page 2)

Task Sheet:
https://drive.google.com/file/d/1UZ_Xy1bvhFchLr7K9hv70IuUHHwjAckf/view?usp=sharing

[Email Signature]

Figure B.8: Invitation to unmoderated usability sessions

Department of Computer Science
t: (807) 343-8310
e: vmago@lakeheadu.ca

Dear Potential Participant,

As a part of a study investigating different survey modes we are conducting a web-based survey on a simulation platform. The purpose of this is to determine whether there is a difference in quality of response and data for moderated and unmoderated usability testing. We are hopeful that this data will aid future researchers in their development of simulation software development.

On behalf of myself, and the rest of our team members, I would like to invite you to participate in the web-based study. By taking part in this endeavour you will be contributing to future developments in simulation software algorithms and technology as well as standing a chance to win a $20 Amazon gift card.

We would greatly appreciate your honest feedback regarding your experience with our software so that we can make improvements to our system. It is important to mention that whether you join this survey or not your academic status or relationship with the Department and Lakehead University will not be affected. Additionally, since this is a volunteer-based survey there will be no academic credit for your participation.

The creation of the simulation scenario and a subsequent survey will take approximately one hour of your time. First, you will be asked to watch a pre-recorded tutorial of how to use our software and then you will be asked to attempt to create a simulation using our website. Following this you will be asked to complete a short usability questionnaire.

Once you have read the information letter and the consent form provided on this page you can start your usability session by following the steps on the task sheet.

Task Sheet:
https://docs.google.com/document/d/1tM7DolhFNwxEBLskLpEHuOrtx-tiN5PO0XNBVmgqspI/edit?usp=sharing

Link to the Information Letter:
https://docs.google.com/document/d/1dD35SEYSaIIQpX__xMFUNh5DNfetm3bJ_pzdbzB6TDo/edit?usp=sharing

Link to the Consent Form:
https://docs.google.com/document/d/1OeJcE-bLY8WLTxXs5V_sWRbb_CE9zF4bnz-5rskTEyk/edit?usp=sharing

Figure B.9: Task sheet for moderated sessions (page 1)

HOMVIZ Survey

Website address: https://homviz.datalab.science/register

<table>
<tr><td colspan="2">Create a new model:

Once you are on the create simulation page you can start making a new simulation model. In the create simulation page you will see a stepper window (tabs) that consists of 5 steps. Below we provided all the instructions on how to fill up every step on that page. Follow the instructions one by one and click on the next button to move to the next page on the stepper.</td></tr>
<tr><td>Step 1: Name and location</td><td>1. Choose any name for your simulation and select Montreal as the city.
2. Move to the next step.</td></tr>
<tr><td>Step 2: Population Group</td><td>The information below shows you which population types to select. From the drop-down list on the page select the following types and once added enter the population counts for each.

Type: under 30, homeless less than 1 year, male
Count: 100

Type: under 30, homeless less than 1 year, female
Count: 100

Type: greater than 50 years, homeless less than 1 year, male
Count: 50</td></tr>
<tr><td>Step 3: Resources</td><td>From the provided drop-down list select the following. Once selected use the control provided on the system page to add sub-elements or enter parameters.

● Select and add shelter. Once added, divide the shelter into <u>two sub-resources</u>:
○ Men’s’ homeless shelters.
■ Only allow men to enter.</td></tr>
</table>

Figure B.10: Task sheet for moderated sessions (page 2)

	■ The initial population for "under 30, homeless less than 1 year, male" is 100. ■ The initial population for "greater than 50 years, homeless less than 1 year, male" is 25. ■ The maximum length of stay is 8 weeks. ■ Capacity is set initially at 200. ○ Women's homeless shelters. ■ Only allow women to enter. ■ The initial population for "under 30, homeless less than 1 year, female" is 50. ■ The maximum length of stay is 7 weeks. ■ Capacity is set initially at 100. ● Select and add **Transitional housing** ○ Allows all population types ○ The initial population is 0 for all population types ○ The maximum length of stay is 52 weeks ○ Capacity is set initially at 200 ● Select and add **Addiction/rehab** centers ○ Allow all population types ○ The initial population is 0 for all population types ○ The maximum length of stay is 24 weeks ○ The capacity is infinite
Step 4: Living situations	From the provided drop-down list select the following. Once selected use the control provided on the system page to enter parameters. ● Select and add **Hidden homeless** ○ This refers mainly to people who are staying with others for a limited time period ("couch-surfing"). ○ Allows all population types ○ The initial population count for "under 30, homeless less than 1 year, female" is 50, and 0 for other types. ● Select and add **Not homeless** ○ This refers to people who are not currently homeless but have been in the recent past ○ Allows all population types ○ The initial population count for "greater than 50 years, homeless less than 1 year, male" is 25, and 0 for others.

Figure B.11: Task sheet for moderated sessions (page 3)

Step 5: Parameters	1. Set number of weeks to 52 2. Set number of simulation to 2 3. Create a simulation

Questionnaire Link: https://homviz.datalab.science/questionnaire

Thank you for your participation.

Figure B.12: Task sheet for unmoderated sessions (page 1)

HOMVIZ Survey

Please follow the instructions given on this page step by step. If followed properly it will take you less than 60 minutes to complete the survey.

1. Navigate to the HOMVIZ Registration Page and create an account.

2. In order to complete the task on our website, you must be familiar with our application. The simulation software is difficult to use without prior knowledge. Please watch the video tutorial that we have made for you to become familiar with our software. https://homviz.datalab.science/tutorial-video

3. Once you have watched the tutorial video, navigate to the homepage, and follow the instructions to create a simulation model:

   a. On the homepage, https://homviz.datalab.science/ Click on the "Create a New Simulation" link.

   b. You are almost done. Follow the steps below to create your own simulation model:

Create a new model:

Once you are on the create simulation page you can start making a new simulation model. In the create simulation page you will see a stepper window (tabs) that consists of 5 steps. Below we provided all the instructions on how to fill up every step on that page. Follow the instructions one by one and click on the next button to move to the next page on the stepper.

Figure B.13: Task sheet for unmoderated sessions (page 2)

Step 1: Name and location	1. Choose any name for your simulation and select Montreal as the city. 2. Move to the next step.
Step 2: Population Group	The information below shows you which population types to select. From the drop-down list on the page select the following types and once added enter the population counts for each. **Type:** under 30, homeless less than 1 year, male **Count:** 100 **Type:** under 30, homeless less than 1 year, female **Count:** 100 **Type:** greater than 50 years, homeless less than 1 year, male **Count:** 50
Step 3: Resources	From the provided drop-down list select the following. Once selected use the control provided on the system page to add sub-elements or enter parameters. • Select and add shelter. Once added, divide the shelter into two sub-resources: ○ Men's' homeless shelters. ■ Only allow men to enter. ■ The initial population for "under 30, homeless less than 1 year, male" is 100. ■ The initial population for "greater than 50 years, homeless less than 1 year, male" is 25. ■ The maximum length of stay is 8 weeks. ■ Capacity is set initially at 200. ○ Women's homeless shelters. ■ Only allow women to enter. ■ The initial population for "under 30, homeless less than 1 year, female" is 50. ■ The maximum length of stay is 7 weeks. ■ Capacity is set initially at 100. • Select and add **Transitional housing** ○ Allows all population types ○ The initial population is 0 for all population types ○ The maximum length of stay is 52 weeks ○ Capacity is set initially at 200

Figure B.14: Task sheet for unmoderated sessions (page 3)

	• Select and add **Addiction/rehab** centers
	  ○ Allow all population types
	  ○ The initial population is 0 for all population types
	  ○ The maximum length of stay is 24 weeks
	  ○ The capacity is infinite

Step 4: Living situations	From the provided drop-down list select the following. Once selected use the control provided on the system page to enter parameters.
	
	• Select and add **Hidden homeless**
	  ○ This refers mainly to people who are staying with others for a limited time period ("couch-surfing").
	  ○ Allows all population types
	  ○ The initial population count for "under 30, homeless less than 1 year, female" is 50, and 0 for other types.
	
	• Select and add **Not homeless**
	  ○ This refers to people who are not currently homeless but have been in the recent past
	  ○ Allows all population types
	  ○ The initial population count for "greater than 50 years, homeless less than 1 year, male" is 25, and 0 for others.

Step 5: Parameters	1. Set number of weeks to 52 2. Set number of simulation to 2 3. Create a simulation

4. Wait for the simulation to finish processing. Click on the results button and view the results of your simulation.

5. Final step. Visit the questionnaire URL, answer the questions, and submit the form.
   https://homviz.datalab.science/questionnaire

Thank you for your participation.

Appendix C

Abbreviations

Abbreviation	Description
UI	User interface
GUI	Graphical user interface
AI	Artificial intelligence
CR	Careless responding
TAM	Technology acceptance model
UTAUT	Unified Theory of Acceptance and Use of Technology
SUS	System Usability Scale
PSSUQ	Post Study System Usability Questionnaire
CSUQ	Computer System Usability Questionnaire
ASQ	After-Scenario Questionnaire
CUQ	Computer Usage Questionnaire
SD	Standard deviation
IoT	Internet of things
COP	Combinatorial Optimization Problem
NP-hardness	Non-deterministic polynomial-time hardness
LP	Linear programming
VP	Vulnerable person